Therefore, using (2.27), we have

$$D_f(x,x^k) - D_f(x,x^{k+1}) = \sum_{i=1}^{m} \lambda_i D_f(x^{k+1}, P_{H_i(\alpha,x^k)}(x^k)) \tag{2.29}$$

$$+ \sum_{i=1}^{m} \lambda_i \left(D_f(x,x^k) - D_f(x, P_{H_i(\alpha,x^k)}(x^k)) \right). \tag{2.30}$$

But, by Theorem 2.0.7,

$$D_f(P_{H_i(\alpha,x^k)}(x^k), x^k) \le D_f(x,x^k) - D_f(x, P_{H_i(\alpha,x^k)}(x^k))$$

for every $x \in C \cap \bar{S}$ and so (2.29) becomes

$$D_f(x,x^k) - D_f(x,x^{k+1}) \ge \sum_{i=1}^{m} \lambda_i (D_f(x^{k+1}, P_{H_i(\alpha,x^k)}(x^k)) + D_f(P_{H_i(\alpha,x^k)}(x^k), x^k)).$$

Hence, since the weights λ_i are positive,

$$D_f(x,x^k) - D_f(x,x^{k+1}) \ge \sum_{i=1}^{m} \lambda_i D_f(P_{H_i(\alpha,x^k)}(x^k), x^k) \tag{2.31}$$

and (2.28) follows since D_f is nonnegative. In the literature, this means that the sequence $\{x^k\}$ is D_f-Fejér monotone with respect to $C \cap \bar{S}$ and implies that $\{x^k\}$ is bounded. This is because a repeated use of (2.28) gives $x^k \in L(x, \alpha_0)$ for all $k \ge 0$ with $\alpha_0 = D_f(x, x^0)$ and so $\{x^k\}$ is bounded by the definition of the Bregman function (see Properties 2.2.3(iv)). □

Several important consequences can be deduced from Proposition 2.2.4. We begin with the next corrollary.

Corollary 2.2.5. *The sequence generated by the algorithm defined by (2.26) is such that for every $y \in C$*

$$\lim_{k\to\infty} D_f(y, x^k) = \theta,$$

for some nonnegative θ and for $i \in I := \{1, \dots, m\}$,

$$\lim_{k\to\infty} D_f(P_{H_i(\alpha,x^k)}(x^k), x^k) = 0. \tag{2.32}$$

Proof. By (2.28), for every $y \in C \cap \bar{S}$, the sequence $\{D_f(y, x^k)\}$ is monotonically decreasing, bounded below by zero and therefore convergent. Hence, there exists a nonnegative θ such that $\lim_{k\to\infty} D_f(y, x^k) = \theta$. Also, as observed in the proof of Proposition 2.2.4, the sequence $\{x^k\}$ is bounded and so the left hand side of (2.31) tends to zero. Therefore $\lim_{k\to\infty} D_f(P_{H_i(\alpha,x^k)}(x^k), x^k) = 0$ since the λ_i's are positive and bounded away from zero. □

Since the sequence generated by the algorithm defined by (2.26) is bounded, it has a limit point. In the next proposition, we show that if a limit point exists and it belongs to C then it is the limit of the sequence.

Proposition 2.2.6. *Any limit point $x^* \in C$ of the bounded sequence $\{x^k\}$ of the algorithm defined by (2.26) is the limit of the entire sequence.*

Proof. Suppose $x^* \in C$ is a limit point of $\{x^k\}$ and that x^{**} is another limit point of $\{x^k\}$. That is

$$\lim_{k\in N_1;k\to\infty} x^k = x^* \text{ and } \lim_{k\in N_2;k\to\infty} x^k = x^{**},$$

where N_1 and N_2 are two different infinite subsets of $N = \{0, 1, 2, \dots, \}$. Then, since f is zone consistent with respect to the C_i's and $x^k \in S$ for all $k \geq 0$, $x^* \in \bar{S}$. Therefore, from (2.28), $\lim_{k\to\infty} D_f(x^*, x^k)$ exists. Applying Properties 2.2.3 (v) in the definition of Bregman function to the sequence $\{x^k\}_{k\geq 0, k\in N_1}$, we have $\lim_{k\in N_1;k\to\infty} D_f(x^*, x^k) = 0$ and so by (2.28) $\lim_{k\to\infty} D_f(x^*, x^k) = 0$ for all $k \geq 0$ which also holds for the sequence $\{x^k\}_{k\geq 0, k\in N_2}$. Therefore using Properties 2.2.3 (vi), we have $x^* = x^{**}$. □

Before the next step for the convergence proof, we need a condition that must be satisfied by the separating hyperplanes in order to guarantee convergence. This condition induces the following definition.

Definition 2.2.7. *For a given closed convex set C and a point $x \notin C$ we say that H is a δ separating hyperplane for C and x if H lies between C and $B(x, \delta d(x, P_C x))$, where $d(\cdot,\cdot)$ stands for the Euclidean distance and $\delta \in (0, 1)$ and $B(x, r)$ is the ball with center at x and radius r.* ◊

It is clear that this definition is equivalent to saying that for every $x \in \mathbb{R}^n$,

$$\|P_H(x) - x\| \geq \delta \|P_C(x) - x\|, \tag{2.33}$$

where $P_C(x)$ and $P_H(x)$ are the Bregman projections of x onto C and the separating hyperplane H respectively. In other words, if the Euclidean distance between x and its projection onto the hyperplane tends to zero, the same is valid for the Euclidean distance between x and its projection onto the associated convex set.

Therefore our assumption for the separating hyperplanes $H_i(\alpha, x)$ in (2.26) for $i = 1, \ldots, m$ will be

Assumption 2.2.8. *For the algorithm defined by (2.26), and for $i \in I := 1, \ldots, m$ and for all $x \in S$ which are not in C_i, the inequality (2.33) holds with $H_i(\alpha, x)$ defined in (2.20) with respect to C_i for $\alpha \in (0, 1)$.* ♢

Given that the algorithm defined by (2.26) satisfies Assumption 2.2.8, we can prove that every limit point of the sequence $\{x^k\}$ generated by (2.26) belong to C. The following theorem justifies this statement.

Theorem 2.2.9. *The whole sequence of the algorithm defined by (2.26) converges to a point in C.*

Proof. Using (2.32) and (2.33), we have

$$\lim_{k\to\infty} D_f(P_{C_i}(x^k), x^k) = 0, \ \forall i \in I := \{1, \ldots, m\}. \tag{2.34}$$

Suppose x^* is a limit point of the sequence $\{x^k\}$. The limit point x^* exists because $\{x^k\}$ is bounded by Proposition 2.2.4. This means that there exists a subsequence $\{x^{k_l}\}$ such that $\lim_{l\to\infty} x^{k_l} = x^*$, and using (2.34)

$$\lim_{l\to\infty} D_f(P_{C_i}(x^{k_l}), x^{k_l}) = 0 \text{ for all } i \in I. \tag{2.35}$$

Now, by Theorem 2.0.7,

$$D_f(x, P_{H_i(\alpha,x^k)}(x^k)) \leq D_f(x, x^k) - D_f(P_{H_i(\alpha,x^k)}(x^k), x^k)$$

for every $x \in C \cap \bar{S}$. But by (2.32) and (2.28), $\{D_f(P_{H_i(\alpha,x^k)}(x^k), x^k)\}$ and $\{D_f(x, x^k)\}$ are bounded for any $x \in C \cap \bar{S}$ and so $\{D_f(x, P_{H_i(\alpha,x^k)}(x^k))\}$ is bounded. Therefore by Properties 2.2.3 (iv), $\{P_{H_i(\alpha,x^k)}(x^k)\}$ is bounded. The boundedness of $\{P_{C_i}(x^k)\}$ follows from Assumption 2.2.8. Therefore, using Properties 2.2.3 (vi), $\lim_{l\to\infty} P_{C_i}(x^{k_l}) = x^*$ for each $i \in I$. Hence $x^* \in C$. □

2.2.2 A general underrelaxed entropy projection method

In this section, we derive a general block iterative algorithm with underrelaxed entropy projections using projections onto separating hyperplanes. We derive this algorithm for the solution of a linear system of equations $Ax = b$.

The $x \log x$ entropy function is a Bregman function with zone $S = \text{Int}\mathbb{R}^n_+$, see page 33 of [32]. Therefore, by Lemma 1.6.10, the Bregman projection $P_{H_i(\alpha,x)}(x)$ of x onto $H(\alpha, x)$ satisfies the equation

$$\nabla f(P_{H_i(\alpha,x)}(x)) = \nabla f(x) + \theta(\nabla f(x) - \nabla f(P_{C_i}(x))), \tag{2.36}$$

where θ is the parameter associated with the projection of x onto $H(\alpha, x)$. For $f(x) = \sum_{j=1}^n (x_j \log x_j)$, the gradient of the jth component is $\nabla f(x)_j = 1 + \log x_j$. Thus using (2.36), we have

$$1 + \log(P_{H_i(\alpha,x^k)}(x^k))_j = 1 + \log x_j^k + \theta_i^k(\log x_j^k - \log(P_{C_i}(x^k))_j)$$

which simplifies to

$$\log(P_{H_i(\alpha,x^k)}(x^k))_j = \log x_j^k + \theta_i^k \log\left(\frac{x_j^k}{(P_{C_i}(x^k))_j}\right) \tag{2.37}$$

and by the definition of the general Bregman method given in (2.26) for $p = 1$,

$$\begin{aligned}
\nabla f(x^{k+1})_j &= \sum_{i=1}^m \lambda_i \nabla f(P_{H_i(\alpha,x^k)}(x^k))_j \\
1 + \log(x^{k+1})_j &= \sum_{i=1}^m \lambda_i (1 + \log(P_{H_i(\alpha,x^k)}(x^k))_j \\
\log(x^{k+1})_j &= \sum_{i=1}^m \lambda_i \log(P_{H_i(\alpha,x^k)}(x^k))_j \text{ for } j = 1, \ldots, n,\ k \geq 0.
\end{aligned}$$

Therefore, using (2.37), we have

$$\begin{aligned}\log x_j^{k+1} &= \sum_{i=1}^{m} \lambda_i \left(\log x_j^k + \theta_i^k \log \left(\frac{x_j^k}{(P_{C_i}(x^k))_j} \right) \right) \\ &= \log x_j^k + \sum_{i=1}^{m} \log \left(\frac{x_j^k}{(P_{C_i}(x^k))_j} \right)^{\lambda_i \theta_i^k} .\end{aligned}$$

Therefore, the iterative step becomes

$$x_j^{k+1} = x_j^k \prod_{i=1}^{m} \left(\frac{x_j^k}{(P_{C_i}(x^k))_j} \right)^{\lambda_i \theta_i^k} . \tag{2.38}$$

But $H_i = \{x \mid \langle a^i, x \rangle = b_i\}$, $a^i \neq 0$, for $i = 1, 2, \ldots, m$, and the jth component of the normal a_j^i is

$$\begin{aligned}a_j^i &= \nabla f(x^k)_j - \nabla f(P_{C_i}(x^k))_j = 1 + \log x_j^k - (1 + \log(P_{C_i}(x^k))_j) \\ &= \log \left(\frac{x_j^k}{(P_{C_i}(x^k))_j} \right).\end{aligned}$$

This implies that

$$\exp a_j^i = \frac{x_j^k}{(P_{C_i}(x^k))_j}.$$

Therefore, using (2.38), the iterative method becomes

$$x_j^{k+1} = x_j^k \prod_{i=1}^{m} (\exp a_j^i)^{\lambda_i \theta_i^k} = x_j^k \prod_{i=1}^{m} \exp\left(a_j^i \lambda_i \theta_i^k\right) \text{ for } j = 1, \ldots, n, \; k \geq 0$$

and if we replace the θ_i^k's with $d_i^k = \log \frac{b_i}{\langle a^i, x^k \rangle}$ for all i and $k \geq 0$ then the resulting formula resembles the iterative step formula of the block-iterative MART algorithm of Censor and Segman [31]. However if one replaces the θ_i^k's with $c_k = \alpha \log \frac{b_i}{\langle a^i, x^k \rangle}$, $0 < \alpha \leq 1$, for all i and $k \geq 0$, then the resulting formula resembles the iterative step formula of the underrelaxed MART algorithm in [23].

2.3 An application for general convex sets

When an exact Bregman projection onto a closed convex set is too costly to compute, we can use the algorithm defined by (2.26)when $p = 1$ and the separating hyperplane defined by the approximation of the function that defines the convex set. That is, suppose that the convex set C_i is defined by

$$C_i = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0\}, \tag{2.39}$$

where $g_i : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function. We follow closely the definition of the approximating hyperplane on page 151 of [61], i.e., a separating hyperplane, $H_i(\alpha, x^k)$ for $\alpha \in (0, 1)$ and $x^k \in S = \text{Int}(\text{dom}g)$, between the current iterate x^k and the current constraint C_i if $x^k \notin C_i$ as

$$H_i(\alpha, x^k) = \{x \in \mathbb{R}^n \mid \alpha g_i(x^k) + \nabla g_i(x^k)^T(x - x^k) = 0\}, \tag{2.40}$$

and $H_i(\alpha, x^k) = \mathbb{R}^n$ if $x^k \in C_i$. We assume $C = \cap_{i=1}^m C_i \neq \emptyset$.

For the algorithm defined by (2.26), $P_{H_i(\alpha,x^k)}(x^k)$ is the Bregman projection of x^k onto $H_i(\alpha, x^k)$. This means that for a function $f \in B(S)$ with $S = \text{Int}(\text{dom}f)$ which is strongly zone consistent with respect to the hyperplane H_i, the projection solves the equations given by

$$\nabla f(P_{H_i(\alpha,x^k)}(x^k)) = \nabla f(x^k) + \theta_i^k \nabla g_i(x^k), \tag{2.41}$$

$$\alpha g_i(x^k) + \nabla g_i(x^k)^T(P_{H_i(\alpha,x^k)}(x^k) - x^k) = 0 \tag{2.42}$$

for the projection parameter θ_i^k and $P_{H_i(\alpha,x^k)}(x^k)$.

Now from (2.41), we have

$$\langle \nabla f(P_{H_i(\alpha,x^k)}(x^k)) - \nabla f(x^k), P_{H_i(\alpha,x^k)}(x^k) - x^k \rangle = \theta_i^k \langle \nabla g_i(x^k), P_{H_i(\alpha,x^k)}(x^k) - x^k \rangle.$$

Therefore, using (2.42), we have

$$\langle \nabla f(P_{H_i(\alpha,x^k)}(x^k)) - \nabla f(x^k), P_{H_i(\alpha,x^k)}(x^k) - x^k \rangle = -\alpha \theta_i^k g_i(x^k).$$

Thus if $x^k \notin C_i$ or $g_i(x^k) > 0$ then

$$\theta_i^k = \frac{\langle \nabla f(P_{H_i(\alpha,x^k)}(x^k)) - \nabla f(x^k), P_{H_i(\alpha,x^k)}(x^k) - x^k \rangle}{-\alpha g_i(x^k)}$$

and (2.41) becomes

$$\nabla f(P_{H_i(\alpha,x^k)}(x^k)) = \nabla f(x^k) - \frac{\langle \nabla f(P_{H_i(\alpha,x^k)}(x^k)) - \nabla f(x^k), P_{H_i(\alpha,x^k)}(x^k) - x^k \rangle}{\alpha g_i(x^k)} \nabla g_i(x^k) \tag{2.43}$$

and if $x^k \in C_i$ or $g_i(x^k) \leq 0$ then $\nabla f(P_{H_i(\alpha,x^k)}(x^k)) = \nabla f(x^k)$.

Observe that if $\nabla g_i(x^k) = 0$ then g_i takes its minimal value at x^k, implying by the non-emptyness of C that $g_i(x^k) \leq 0$, so that $P_{H_i(\alpha,x^k)}(x^k) = x^k$ and so $\nabla f(P_{H_i(\alpha,x^k)}(x^k)) = \nabla f(x^k)$.

In order that the algorithm defined by (2.26) converges to $x^* \in C$ for the general convex set, we make the following assumption on the gradient of the function g_i for $i = 1, \ldots, m$. With this assumption, i.e., Assumption 2.3.1, Proposition 2.3.2 guarantees that $x^* \in C$ and ensures that Assumption 2.2.8 is satisfied.

Assumption 2.3.1. *For each of the functions $\{g_i\}$ used in (2.39), there exists $\epsilon > 0$ such that $\|\nabla g_i(x^k)\| \geq \epsilon$.*

Proposition 2.3.2. *On the basis of Assumption 2.3.1, $\lim_{k\to\infty} g_i(x^k) = 0$ and $\lim_{k\to\infty} x^k \in C_i$ for each $i \in \{1, \ldots, m\}$.*

Proof. If $\|\nabla g_i(x^k)\|$ is bounded away from zero as in Assumption 2.3.1 then (2.32) implies that $P_{H_i(\alpha,x^k)}(x^k) - x^k$ tends to zero as $k \to \infty$, and so (2.42) implies that $\alpha \lim_{k\to\infty} g_i(x^k) = 0$. Therefore, $\lim_{k\to\infty} x^k \in C_i$ for each i since the g_i's are continuous and $\alpha \in (0, 1)$. This further implies that $\lim_{k\to\infty}(P_{C_i}(x^k) - x^k) = 0$ which satisfies condition (2.33). □

Remark 2.3.3. It must be noted that, for the general convex sets, Assumption 2.3.1 is sufficient for the proof of Theorem 2.2.9. Assumption 2.2.8 is thus redundant.

The proof of Theorem 2.2.9 may therefore take the following simple form:

Proof. By Proposition 2.2.4, $\{x^k\}$ is bounded and so there exists a subsequence $\{x^{k_l}\}$ such that $\lim_{l\to\infty} x^{k_l} = x^*$. Therefore, by Proposition 2.3.2, $x^* \in C_i$ for each $i \in \{1, \ldots, m\}$. Thus $x^* \in C$ and hence by Proposition 2.2.6, x^* is the limit of the entire sequence. □

2.4 Linear equality constraints

In the case of linear equality constraints, given by a linear system of equations

$$Ax = b,$$

where $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, if for some function $f \in B(S)$ which is strongly zone consistent with respect to the hyperplane $H_i := \{x \in \mathbb{R}^n \mid \langle a^i, x \rangle = b_i\}$, $a^i \neq 0$, for each $i \in \{1, \ldots, m\}$, the starting point of the algorithm defined by (2.26) is given by

$$\nabla f(x^0) = -A^T z^0$$

for a given $z^0 \in \mathbb{R}^m_+$ (observe that the obvious choice is $z^0 = 0$ and x^0 would be the unconstrained minimum of f) then we have the iterative formula given by

$$\nabla f(x^{k+1}) = \sum_{i=1}^{m} \lambda_i \nabla f(P_{H_i(\alpha, x^k)}(x^k))$$

where, from (1.34), $\nabla f(x^k) = -A^T z^k$ for $k \geq 0$ and so, in the limit, say x^*, $\nabla f(x^*) = -A^T z^*$, for some z^*. Therefore the Kuhn-Tucker conditions are satisfied for the optimization problem

$$\min f(x) \text{ subject to } Ax = b. \tag{2.44}$$

In this case, algorithm (2.26) solves (2.44) when $p = 1$.

2.5 A Conjecture for the strongly underrelaxed case

In the general case, where the constraints are not linear equalities, just the sequence of Bregman projections defined by algorithm (2.26) does not guarantee convergence to the solution of a minimization problem, say

$$\min f(x) \text{ subject to } \langle a^i, x \rangle \leq b_i$$

for the function $f \in B(S)$ with $S = \text{Intdom} f$ and $i = 1, \ldots, m$, but converges to the solution of only the convex feasibility problem, unless dual variables are updated as in Chapter 3 for linear inequalities. So, what happens in the general case? Our conjecture is that in the purely sequential Bregman algorithm, when the relaxation parameters tend to zero, the sequence generated by algorithm (2.26) tends to the solution of the optimization problem $\lim_{x\in\mathbb{R}^n} D_f(Ax, b)$ in the general case.

Chapter 3

Block Bregman methods for inequality constraints

In this chapter, we present a new simultaneous version of the Bregman method for linear constraints with corresponding convergence results.

3.1 The problem

We recall the problem of Subsection 1.7.1,

$$\begin{aligned} &\min f(x), \\ \text{subject to } &\langle a^i, x\rangle \le b_i, \qquad i \in I := \{1, 2, \dots, m\}, \\ &x \in \bar{S}, \end{aligned}$$

where $H_i := \{x \in \mathbb{R}^n \mid \langle a^i, x\rangle = b_i\}$, $C_i := \{x \in \mathbb{R}^n \mid \langle a^i, x\rangle \le b_i\}$; $C = \cap_{i=1}^m C_i$ and $C \cap \bar{S} \neq \emptyset$.

A is an $m \times n$ matrix whose ith row is a^i, and $b \in \mathbb{R}^m$, $a^i \neq 0$ for all $i \in I$. $S = \mathrm{Int}(\mathrm{dom} f)$ and $f \in B(S)$ is essentially smooth and strongly zone consistent with respect to every H_i.

3.1.1 Simultaneous under-relaxed Bregman's algorithm for linear inequality constraints

We present the simultaneous version of Algorithm 1.7.2 for solving problem (1.43)-(1.45).

Algorithm 3.1.1. Simultaneous under-relaxed Bregman's algorithm for linear inequalities

(i) **Initialization** $x^0 \in S$ is such that for an arbitrary $z^0 \in \mathbb{R}^m_+$,

$$\nabla f(x^0) = -A^T z^0. \tag{3.1}$$

(ii) **Iterative Step** Given x^k and z^k, calculate x^{k+1} and z^{k+1} from

$$\nabla f(x^{k+1}) = \nabla f(x^k) + \sum_{i \in I_{t(k)}} \lambda_i^k c_i^k a^i, \tag{3.2}$$

$$z^{k+1} = z^k - \sum_{i \in I_{t(k)}} \lambda_i^k c_i^k e^i \tag{3.3}$$

with

$$c_i^k = \begin{cases} \min\left(\frac{z_i^k}{\lambda_i^k}, \theta_i^k\right) & \text{if} \quad i \in I_{t(k)}, \\ 0 & \text{if} \quad i \notin I_{t(k)}, \end{cases} \tag{3.4}$$

where $\theta_i^k = \pi_{H(k)_i}(x^k)$, and there exists $\bar{\epsilon} > 0$ such that, for the positive weights λ_i^k with $\sum_{i \in I_{t(k)}} \lambda_i^k = 1$ for all $k \geq 0$, $\lambda_i^k \geq \bar{\epsilon}$ for $i \in I_{t(k)}$ and $\lambda_i^k = 0$ for $i \notin I_{t(k)}$.

(iii) $H(k)_i$ is parallel to $H_i = \{x \in \mathbb{R}^n \mid \langle a^i, x \rangle = b_i\}$ for all $i \in I_{t(k)}$ and $k \geq 0$. $H(k)_i := \{x \in \mathbb{R}^n \mid \langle a^i, x \rangle = \alpha^k b_i + (1 - \alpha^k)\langle a^i, x^k \rangle\}$, where $\{\alpha^k\}$ such that $\epsilon \leq \alpha^k \leq 1$ for $\epsilon > 0$ is a sequence of relaxation parameters. We adopt the concept of relaxation highlighted in Subsection 1.7.2 and implemented in [41].

(iv) The sequence $\{t(k)\}_{k=0}^{\infty}$ is almost cyclic on the index set $\{1, 2, \ldots, M\}$, where M is the number of blocks. The index set $I := \{1, 2, \ldots, m\}$ of the m constraints has been partitioned into M nonempty disjoint blocks such that $I = I_1 \cup \ldots \cup I_M$. ♢

In order to justify the proof of Lemma 3.1.4 and the proofs of the propositions leading to the proof of the convergence theorem, Theorem 3.2.1, let P_i and Q_i be Bregman projections onto

$$C(k)_i = \{x \in \mathbb{R}^n \mid \langle a^i, x\rangle \leq \alpha^k b_i + (1-\alpha^k)\langle a^i, x^k\rangle\} \tag{3.5}$$

and $H(k)_i$ respectively so that for any function $f \in B(S)$ which is strongly zone consistent with respect to H_i for $i \in I_{t(k)}$ and $x^k \in S$ with $\theta_i^k = \pi_{H(k)_i}(x^k)$,

$$\nabla f(P_i x^k) = \nabla f(x^k) + \min\{0, \theta_i^k\} a^i, \tag{3.6}$$

$$\nabla f(Q_i x^k) = \nabla f(x^k) + \theta_i^k a^i. \tag{3.7}$$

Define

$$\nabla f(w_i^k) = \nabla f(x^k) + c_i^k a^i \;\text{ for } i \in I_{t(k)}. \tag{3.8}$$

The w_i^k's are 'modified relaxed Bregman projections' onto the H_i's (see page 233 of [62]) such that the next iterate x^{k+1} satisfies the equation $\nabla f(x^{k+1}) = \sum_{i\in I_{t(k)}} \lambda_i^k \nabla f(w_i^k)$.

This is because, multiplying (3.8) by λ_i^k for $i \in I_{t(k)}$ with $\sum_{i\in I_{t(k)}} \lambda_i^k = 1$ and summing over $i \in I_{t(k)}$, we have

$$\sum_{i\in I_{t(k)}} \lambda_i^k \nabla f(w_i^k) = \nabla f(x^k) + \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k a^i.$$

Therefore, from iterative step (3.2),

$$\nabla f(x^{k+1}) = \sum_{i\in I_{t(k)}} \lambda_i^k \nabla f(w_i^k). \tag{3.9}$$

Since the applicability of Algorithm 3.1.1 depends on the ability to invert the gradient ∇f explicitly, we assume that the Bregman function f used in Algorithm 3.1.1 satisfies Assumption 2.2.2.

3.1.2 Preliminary results

The following lemma establishes the non-negativity of the dual variables.

Lemma 3.1.2. *For any $k \geq 0$, $z_i^k \geq 0$ for $i \in I_{t(k)}$.*

Proof. By initialization, $z^0 \in \mathbb{R}^m_+$. By induction, assume $z^k \in \mathbb{R}^m_+$. If $i \notin I_{t(k)}$ then $z_i^{k+1} = z_i^k \geq 0$ and if $i \in I_{t(k)}$ then using (3.4), $z_i^{k+1} = z_i^k - \lambda_i^k c_i^k \geq z_i^k - z_i^k = 0$. Hence $z^{k+1} \in \mathbb{R}^m_+$. □

Lemma 3.1.3. $\nabla f(x^k) = -A^T z^k$ *for any* $k \geq 0$.

Proof. We proceed by induction. By (3.1), the result is true when $k = 0$. Assume that the result is true for any $k \geq 0$. Then by (3.2) and (3.3),

$$\nabla f(x^{k+1}) = \nabla f(x^k) + \sum_{i \in I_{t(k)}} \lambda_i^k c_i^k a^i \tag{3.10}$$

$$= -A^T(z^k - \sum_{i \in I_{t(k)}} \lambda_i^k c_i^k e^i) = -A^T z^{k+1}. \tag{3.11}$$

□

The next lemma shows that the w_i^k is a Bregman projection of x^k onto $H(k)_i$ if $x^k \notin C(k)_i$ or a point in the segment between x^k and its projection onto $H(k)_i$ if $x^k \in C(k)_i$.

Lemma 3.1.4. *Let the function* $f \in B(S)$ *be strongly zone consistent with respect to the hyperplane* H_i *for all* $i \in I_{t(k)}$ *and* $k \geq 0$, *and let* w_i^k *satisfy (3.8). Then for the closed and nonempty convex set* C *with* $C \cap \bar{S} \neq \emptyset$, *the following statements hold if* f *satifies Assumption 2.2.2.*

(a) If $x^k \notin C(k)_i$ *then* $w_i^k = P_i x^k = Q_i x^k$ *and* $\theta_i^k = c_i^k < 0$; P_i *and* Q_i *are as defined in (3.6) and (3.7).*

(b) If $x^k \in C(k)_i$ *then* $0 \leq c_i^k \leq \theta_i^k$.

(c) $w_i^k \in C(k)_i$.

(d) $D_f(w_i^k, x^k) \leq D_f(y, x^k) - D_f(y, w_i^k)$ *for all* $y \in C \cap \bar{S}$.

Proof. (a) If $x^k \notin C(k)_i$ then by Lemma 1.6.11 $\theta_i^k < 0$, and since $z_i^k \geq 0$ by Lemma 3.1.2, $c_i^k = \theta_i^k$ by (3.4). Therefore, from the definitions of w_i^k, P_i and Q_i, we have $w_i^k = P_i x^k = Q_i x^k$.

(b) If $x^k \in C(k)_i$ then $\theta_i^k \geq 0$ and since $z_i^k \geq 0$, $0 \leq c_i^k \leq \theta_i^k$.

(c) Clearly $w_i^k \in C(k)_i$ if $c_i^k = \theta_i^k$. Thus, suppose $c_i^k \neq \theta_i^k$ or $c_i^k = z_i^k/\lambda_i^k$, i.e., $c_i^k < \theta_i^k$, and recall that θ_i^k, the parameter associated with the Bregman projection of x^k onto $H(k)_i$ (in what follows we denote this projection by $\bar{x}^{k+1}$) is obtained by solving the system

$$\nabla f(\bar{x}^{k+1}) = \nabla f(x^k) + \theta_i^k a^i, \tag{3.12}$$

$$\langle a^i, \bar{x}^{k+1} \rangle = \alpha^k b_i + (1 - \alpha^k)\langle a^i, x^k \rangle. \tag{3.13}$$

Now, using (3.8) and (3.12), we have

$$\nabla f(w_i^k) - \nabla f(\bar{x}^{k+1}) = (c_i^k - \theta_i^k)a^i$$

and so

$$\langle \nabla f(w_i^k) - \nabla f(\bar{x}^{k+1}), w_i^k - \bar{x}^{k+1} \rangle = \langle (c_i^k - \theta_i^k)a^i, w_i^k - \bar{x}^{k+1} \rangle.$$

But, by Properties 2.2.3 (iii),

$$\langle \nabla f(w_i^k) - \nabla f(\bar{x}^{k+1}), w_i^k - \bar{x}^{k+1} \rangle = D_f(w_i^k, \bar{x}^{k+1}) + D_f(\bar{x}^{k+1}, w_i^k) \geq 0$$

and since $c_i^k < \theta_i^k$, we have

$$\langle a^i, w_i^k - \bar{x}^{k+1} \rangle \leq 0.$$

Therefore

$$\langle a^i, w_i^k \rangle \leq \alpha^k b_i + (1 - \alpha^k)\langle a^i, x^k \rangle$$

since

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

This means that $w_i^k \in C(k)_i$.

(d) The result follows from Theorem 2.0.7.

□

3.2 Convergence results

Next, is the convergence theorem for Algorithm 3.1.1.

Theorem 3.2.1. *Assume the following:*

(i) $f \in B(S)$,

(ii) f *is strongly zone consistent with respect to each* H_i, $i \in I_{t(k)}$,

(iii) $\{t(k)\}_{k=0}^{\infty}$ *is almost cyclic on* $\{1, 2, \dots, M\}$ *with a constant of almost cyclicality* r,

(iv) $C \cap \bar{S} \neq \emptyset$.

Then any sequence $\{x^k\}$ *produced by Algorithm 3.1.1 converges to the point* x^**, which is the solution of (1.43)-(1.45).*

To prove Theorem 3.2.1, we use Propositions 3.2.2 to 3.2.7.

Proposition 3.2.2. *If the assumptions of Theorem 3.2.1 hold and* w_i^k *satisfies (3.8), then*
$w_i^k = P_{\bar{H}(k)_i}(x^k)$ *and* $c_i^k = \pi_{\bar{H}(k)_i}(x^k)$ *where*

$$\bar{H}(k)_i = \{x \in \mathbb{R}^n \mid \langle a^i, x\rangle = \gamma^k b_i + (1-\gamma^k)\langle a^i, x^k\rangle\} \tag{3.14}$$

for some $\gamma^k \in \mathbb{R}$ *such that* $0 \leq \gamma^k \leq \alpha^k$ *for all* $k \geq 0$.

Proof. Note that w_i^k is the Bregman projection of x^k onto the hyperplane $\bar{H}(k)_i = \{\langle a^i, x\rangle = \langle a^i, w_i^k\rangle\}$, which is parallel to $H(k)_i$ and passes through w_i^k. It remains to demonstrate that the right-hand side $\langle a^i, w_i^k\rangle$ has the desired form as in (3.14). It is clear that if $c_i^k = \theta_i^k$ for $i \in I_{t(k)}$ then $\bar{H}(k)_i = H(k)_i$. This is because, by definition, $\bar{H}(k)_i$ is parallel to $H(k)_i$ and lies between x^k and $H(k)_i$, and θ_i^k is the parameter associated with the Bregman projection of x^k onto $H(k)_i$. Therefore $\bar{H}(k)_i$ and $H(k)_i$ coincide if $c_i^k = \theta_i^k$. Hence $\gamma^k = \alpha^k$.

On the other hand, if $c_i^k \neq \theta_i^k$ for $i \in I_{t(k)}$ then, by the definition of c_i^k, $0 \leq \frac{z_i^k}{\lambda_i^k} = c_i^k < \theta_i^k$.

Now consider the following which is possible by Properties 2.2.3 (iii) and (3.8):

$$\begin{aligned} 0 \leq D_f\left(w_i^k, x^k\right) + D_f\left(x^k, w_i^k\right) &= \langle \nabla f\left(w_i^k\right) - \nabla f\left(x^k\right), w_i^k - x^k \rangle, \\ &= \langle c_i^k a^i, w_i^k - x^k \rangle. \end{aligned} \tag{3.15}$$

Therefore if $c_i^k = 0$ for $i \in I_{t(k)}$ then $D_f\left(x^k, w_i^k\right) = 0$ and $D_f\left(w_i^k, x^k\right) = 0$ imply $w_i^k = x^k$, i.e.,$\langle a^i, w_i^k \rangle = \langle a^i, x^k \rangle$ and so we may take $\gamma^k = 0$.

Finally if $0 < c_i^k < \theta_i^k$ then $0 < \pi_{\bar{H}(k)_i}(x^k) < \pi_{H(k)_i}(x^k)$. Hence, by Lemma 1.6.12,

$$\langle a^i, x^k \rangle < \langle a^i, w_i^k \rangle < \alpha^k b_i + (1 - \alpha^k)\langle a^i, x^k \rangle \tag{3.16}$$

for $i \in I_{t(k)}$ and so there is $\gamma^k \in (0, \alpha^k)$ as desired. For more insight, see page 427 of [41] □

Proposition 3.2.3. *If the assumptions of Theorem 3.2.1 hold, λ_i^k and c_i^k are as defined in Algorithm 3.1.1 and w_i^k satisfies (3.8) then for $x^k \in S$,*

$$\sum_{i \in I_{t(k)}} \lambda_i^k \left\{ D_f(x^k, w_i^k) - D_f(x^k, x^{k+1}) \right\}$$

and

$$D_f\left(x^{k+1}, x^k\right) + \sum_{i \in I_{t(k)}} \lambda_i^k c_i^k \left(b_i - \langle a^i, x^{k+1} \rangle\right)$$

are non-negative.

Proof. Using Lemma 2.0.6 with $y = x^k$, $x = w_i^k$ and $z = x^{k+1}$, we have

$$D_f(x^k, w_i^k) = D_f(x^k, x^{k+1}) + D_f(x^{k+1}, w_i^k) - \langle \nabla f(w_i^k) - \nabla f(x^{k+1}), x^k - x^{k+1} \rangle.$$

Now, multiplying the last equation by λ_i^k for $i \in I_{t(k)}$, $\sum_{i \in I_{t(k)}} \lambda_i^k = 1$, and summing over $i \in I_{t(k)}$, we have

$$\sum_{i \in I_{t(k)}} \lambda_i^k D_f(x^k, w_i^k) = D_f(x^k, x^{k+1}) + \sum_{i \in I_{t(k)}} \lambda_i^k D_f(x^{k+1}, w_i^k)$$

since, by (3.9),

$$\begin{aligned}\sum_{i\in I_{t(k)}} \lambda_i^k \langle \nabla f(w_i^k) - \nabla f(x^{k+1}), x^k - x^{k+1}\rangle &= \langle \nabla f(x^{k+1}) - \nabla f(x^{k+1}), x^k - x^{k+1}\rangle \\ &= 0.\end{aligned}$$

Thus

$$\sum_{i\in I_{t(k)}} \lambda_i^k \left\{ D_f(x^k, w_i^k) - D_f(x^k, x^{k+1}) \right\} = \sum_{i\in I_{t(k)}} \lambda_i^k D_f(x^{k+1}, w_i^k) \geq 0. \qquad (3.17)$$

Now let

$$\bar{d}_k = D_f\left(x^{k+1}, x^k\right) + \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k \left(b_i - \langle a^i, x^{k+1}\rangle\right). \qquad (3.18)$$

Then, using Properties 2.2.3 and (3.2), we have

$$\begin{aligned}\bar{d}_k + D_f\left(x^k, x^{k+1}\right) &= D_f\left(x^k, x^{k+1}\right) + D_f\left(x^{k+1}, x^k\right) \\ &\quad + \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k \left(b_i - \langle a^i, x^{k+1}\rangle\right) \\ &= \langle \nabla f\left(x^{k+1}\right) - \nabla f\left(x^k\right), x^{k+1} - x^k\rangle \\ &\quad + \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k \left(b_i - \langle a^i, x^{k+1}\rangle\right) \\ &= \langle \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k a^i, x^{k+1} - x^k\rangle + \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k \left(b_i - \langle a^i, x^{k+1}\rangle\right) \\ &= \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k \left(b_i - \langle a^i, x^k\rangle\right).\end{aligned}$$

But, by Proposition 3.2.2, $\langle a^i, w_i^k\rangle = \gamma^k b_i + (1-\gamma^k)\langle a^i, x^k\rangle$ implies

$$\langle a^i, w_i^k - x^k\rangle = \gamma^k \left(b_i - \langle a^i, x^k\rangle\right). \qquad (3.19)$$

Therefore, from (3.15),

$$0 \leq D_f\left(w_i^k, x^k\right) + D_f\left(x^k, w_i^k\right) = c_i^k \gamma^k \left(b_i - \langle a^i, x^k\rangle\right). \qquad (3.20)$$

Therefore, using (3.20), if $\gamma^k = 0$ then $w_i^k = x^k$, and by (3.8), $w_i^k = x^k$ implies $c_i^k = 0$. Therefore by (3.2), $\nabla f(x^{k+1}) = \nabla f(x^k)$ implies $x^{k+1} = x^k$ if f satisfies Assumption 2.2.2. Hence

$$\bar{d}_k = -D_f\left(x^k, x^{k+1}\right) + \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k \left(b_i - \langle a^i, x^k\rangle\right) = 0 \text{ if } \gamma^k = 0.$$

Next we consider when $\gamma^k > 0$ for $k \geq 0$. Using (3.20), we have

$$\bar{d}_k = \sum_{i \in I_{t(k)}} \lambda_i^k \left\{ -D_f\left(x^k, x^{k+1}\right) + \frac{D_f\left(w_i^k, x^k\right) + D_f\left(x^k, w_i^k\right)}{\gamma^k} \right\}$$

and since γ^k is less or equal to one, we have

$$\bar{d}_k \geq \sum_{i \in I_{t(k)}} \lambda_i^k \left\{ D_f\left(w_i^k, x^k\right) + D_f\left(x^k, w_i^k\right) - D_f\left(x^k, x^{k+1}\right) \right\}.$$

Therefore, using (3.17),

$$\bar{d}_k \geq \sum_{i \in I_{t(k)}} \lambda_i^k \left\{ D_f\left(w_i^k, x^k\right) + D_f\left(x^{k+1}, w_i^k\right) \right\}. \tag{3.21}$$

Thus

$$\bar{d}_k \geq 0 \text{ for all } k \geq 0.$$

□

Proposition 3.2.4. *If the Lagrangian of the minimization problem in (1.43)-(1.45) is* $L(x,z) = f(x) + \langle z, Ax - b \rangle$ *then, for any sequences* $\{x^k\}$ *and* $\{z^k\}$ *produced by Algorithm 3.1.1,*

(i) the sequence $\left\{L\left(x^k, z^k\right)\right\}$ *is nondecreasing and* $\lim_{k\to\infty} L\left(x^k, z^k\right)$ *exists,*

(ii) $\lim_{k\to\infty} D_f\left(w_i^k, x^k\right) = 0$ *and* $\lim_{k\to\infty} D_f\left(x^{k+1}, w_i^k\right) = 0$ *for each* $i \in I_{t(k)}$*,*

(iii) $\{x^k\}$ *is bounded.*

Proof. (i) Define

$$d_k = L\left(x^{k+1}, z^{k+1}\right) - L\left(x^k, z^k\right). \tag{3.22}$$

Then, from the definition of L,

$$\begin{aligned} d_k &= f(x^{k+1}) + \langle z^{k+1}, Ax^{k+1} - b \rangle - \left(f(x^k) + \langle z^k, Ax^k - b \rangle\right) \\ &= f(x^{k+1}) - f(x^k) + \langle z^{k+1}, Ax^{k+1} \rangle - \langle z^{k+1}, b \rangle - \langle z^k, Ax^k \rangle + \langle z^k, b \rangle. \end{aligned}$$

But $\langle z^k, Ax^k\rangle = \langle A^T z^k, x^k\rangle = -\langle \nabla f(x^k), x^k\rangle$.

Therefore

$$\begin{aligned}
d_k &= f(x^{k+1}) - f(x^k) - \langle \nabla f(x^k), x^{k+1} - x^k\rangle + \langle \nabla f(x^k), x^{k+1} - x^k\rangle \\
&\quad -\langle z^{k+1} - z^k, b\rangle - \langle \nabla f(x^{k+1}), x^{k+1}\rangle + \langle \nabla f(x^k), x^k\rangle \\
&= D_f\left(x^{k+1}, x^k\right) - \langle z^{k+1} - z^k, b\rangle + \langle \nabla f(x^k) - \nabla f(x^{k+1}), x^{k+1}\rangle \\
&= D_f\left(x^{k+1}, x^k\right) - \langle \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k a^i, x^{k+1}\rangle + \langle \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k e^i, b\rangle \\
&= D_f\left(x^{k+1}, x^k\right) - \langle \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k a^i, x^{k+1}\rangle + \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k b_i \\
&= D_f\left(x^{k+1}, x^k\right) + \sum_{i\in I_{t(k)}} \lambda_i^k c_i^k \left(b_i - \langle a^i, x^{k+1}\rangle\right).
\end{aligned}$$

Hence, by Proposition 3.2.3 and (3.18), $d_k = \bar{d}_k \geq 0$ and so $\left\{L(x^k, z^k)\right\}$ is non-decreasing.

We now prove the existence of $\lim_{k\to\infty} L\left(x^k, z^k\right)$ by showing that $\left\{L\left(x^k, z^k\right)\right\}$ is bounded from above on $C \cap \bar{S}$ for all $k \geq 0$.

To do this, we choose $z \in C \cap \bar{S}$ and consider:

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

Therefore

$$L\left(x^k, z^k\right) \leq f(z) - D_f(z, x^k) \leq f(z).$$

Hence $\left\{L\left(x^k, z^k\right)\right\}$ is bounded from above and so the limit exists.

(ii) Since $\lim_{k\to\infty} L\left(x^k, z^k\right)$ exists, by (3.22), $\lim_{k\to\infty} d_k = \lim_{k\to\infty} \bar{d}_k = 0$. Therefore, from (3.21),

$$\lim_{k\to\infty} \sum_{i\in I_{t(k)}} \lambda_i^k \left\{D_f\left(w_i^k, x^k\right) + D_f\left(x^{k+1}, w_i^k\right)\right\} = 0 \qquad (3.23)$$

and since $\lambda_i^k \geq \bar{\epsilon} > 0$ for $i \in I_{t(k)}$ and $k \geq 0$, we have

$$\lim_{k\to\infty} D_f\left(w_i^k, x^k\right) = 0 \text{ and } \lim_{k\to\infty} D_f\left(x^{k+1}, w_i^k\right) = 0$$

for each $i \in I_{t(k)}$ and $k \geq 0$.

(iii) Applying $d_k = L\left(x^{k+1}, z^{k+1}\right) - L\left(x^k, z^k\right) \geq 0$ for all $k \geq 0$ recursively to the expression $D_f(z, x^k) \leq f(z) - L\left(x^k, z^k\right)$, we have

$$D_f(z, x^k) \leq f(z) - L\left(x^0, z^0\right) = \alpha,$$

where $\alpha \in \mathbb{R}$. Therefore, from the boundedness of the partial level set $L_2^f(z, \alpha)$, Properties 2.2.3 (iv), we have that $\{x^k\}$ is bounded.

□

Proposition 3.2.5. *Assume that* $\lim_{j\to\infty} x^{k_j} = x^*$. *Fix* q*, a positive integer, and take a sequence* $\{l_j\}$ *with* $l_j \in \{1, 2, \ldots, q\}$. *Then* $\lim_{j\to\infty} w_i^{k_j+l} = x^*$ *for* $i \in I_{t(k_j)}$ *and* $\lim_{j\to\infty} x^{k_j+l_j} = x^*$.

Proof. Consider first the q sequences $\left\{x^{k_j+l}\right\}_{j=0}^{\infty}$ with $1 \leq l \leq q$. Since $\left\{x^{k_j+l}\right\}_{j=0}^{\infty}$ are sub-sequences of the bounded sequence $\{x^k\}_{k=0}^{\infty}$, they are bounded, and from Proposition 3.2.4 (ii), for all $s \in \{0, 1, \ldots, q\}$, $i \in I_{t(k_j)}$ and $j \geq 0$,

$$\lim_{j\to\infty} D_f\left(w_i^{k_j+s}, x^{k_j+s}\right) = 0 \text{ and } \lim_{j\to\infty} D_f\left(x^{k_j+s+1}, w_i^{k_j+s}\right) = 0. \qquad (3.24)$$

From Proposition 3.2.4 (ii) and (iii), $\{D_f(w_i^k, x^k)\}$ is bounded and $\{D_f(x, x^k)\}$ for each $x \in C \cap \bar{S}$ is bounded . Therefore by Lemma 3.1.4 (d), $\{D_f(x, w_i^k)\}$ is bounded for $x \in C \cap \bar{S}$. This means that, by Properties 2.2.3 (iv), $\{w_i^k\}$ is bounded. Therefore, using the first equation on the left hand side of (3.24) and applying recursively Properties 2.2.3 (vi), we have, for all l such that $0 \leq l \leq q$ and $i \in I_{t(k_j)}$,

$$\lim_{j\to\infty} w_i^{k_j+l} = x^*. \qquad (3.25)$$

Hence, using the second equation on the right hand side of (3.24) and applying recursively Properties 2.2.3 (vi), we have that, for all l such that $0 \leq l \leq q$,

$$\lim_{j\to\infty} x^{k_j+l} = x^*.$$

Thus, interlacing these $q+1$ sequences, we can form the sequences

$$\left\{w_i^{k_1}, w_i^{k_1+1}, \ldots, w_i^{k_1+q}, w_i^{k_2}, w_i^{k_2+1}, \ldots, w_i^{k_2+q}, \ldots, w_i^{k_j}, w_i^{k_j+1} \ldots, w_i^{k_j+q}, \ldots\right\}$$

and

$$\{x^{k_1}, x^{k_1+1}, \ldots, x^{k_1+q}, x^{k_2}, x^{k_2+1}, \ldots, x^{k_2+q}, \ldots, x^{k_j}, x^{k_j+1} \ldots, x^{k_j+q}, \ldots\}$$

both converging to x^* with $\{w_i^{k_j+l_j}\}$ and $\{x^{k_j+l_j}\}$ as their respective sub-sequences.

□

Proposition 3.2.6. *All limit points of the sequence $\{x^k\}$ produced by Algorithm 3.1.1 belong to $C \cap \bar{S}$.*

Proof. Let $p \in \{1, 2, \ldots, M\}$ and $l_j \in \{1, 2, \ldots, r\}$, where r is the constant of almost cyclicality of $\{t(k_j)\}_{j=0}^{\infty}$, such that $t(k_j + l_j) = p$. Then, with $q = r$ as in Proposition 3.2.5 and with $i \in I_{t(k_j+l_j)} = I_p$,

$$\lim_{j\to\infty} w_i^{k_j+l_j} = x^* \text{ and } \lim_{j\to\infty} x^{k_j+l_j} = x^*$$

hold and so we can extract sub-sequences $\{x^{s_j}\}_{j=0}^{\infty}$ of $\{x^{k_j+l_j}\}_{j=0}^{\infty}$ such that $x^{s_j} \to x^*$, $\gamma^{s_j} \to \gamma$, $\alpha^{s_j} \to \alpha \geq \epsilon$, and by Lemma 3.1.4 (c) and Proposition 3.2.2, $x^* \in C(s_j)_i$ as $j \to \infty$. Therefore, it follows that for $i \in I_{t(s_j)}$,

$$\langle a^i, x^* \rangle = \gamma b_i + (1 - \gamma)\langle a^i, x^* \rangle \text{ or } \gamma \left(\langle a^i, x^* \rangle - b_i\right) = 0. \tag{3.26}$$

Therefore if $\gamma \neq 0$ then $\langle a^i, x^* \rangle = b_i$ and if $\gamma^{s_j} \to \gamma = 0$ then (3.26) is inconclusive, and moreover $\gamma \neq \alpha$ since $\alpha^{s_j} \to \alpha \geq \epsilon > 0$, and so by Proposition 3.2.2, precisely (3.16), we have

$$\langle a^i, x^* \rangle \leq \alpha b_i + (1 - \alpha)\langle a^i, x^* \rangle \text{ or } 0 \leq \alpha \left(b_i - \langle a^i, x^* \rangle\right)$$

for $i \in I_p$. Therefore $\langle a^i, x^* \rangle \leq b_i$ for $i \in I_p$. Hence, since p is an arbitrary index, we have $Ax^* \leq b$ and so $x^* \in C$. □

Proposition 3.2.7. *For $x \in C$, define $I_1^{t(k)}(x) := \{i \in I_{t(k)} : \langle a^i, x \rangle < b_i\}$ and $I_2^{t(k)}(x) := \{i \in I_{t(k)} : \langle a^i, x \rangle = b_i\}$ for all $k \geq 0$ and assume that $\lim_{j\to\infty} x^{k_j} = x^*$. Then $z_i^{k_j+r+1} = 0$ for all $i \in I_1^{t(k_j)}(x^*)$ and $j \geq 0$, where r is the constant of almost cyclicality.*

Proof. Let

$$\delta = \frac{\epsilon}{4} \min_{i \in I_1^{t(k_j)}(x^*)} \left\{ \frac{(b_i - \langle a^i, x^* \rangle)}{\|a^i\|} \right\} > 0.$$

By Proposition 3.2.5, there exists a natural number J such that

$$\|w_i^{k_j+l} - x^*\| < \delta \text{ for all } l \in \{0, \ldots, r+1\}, \text{ and } i \in I_1^{(t(k_j)+l)}(x^*) \text{ for all } j \geq J. \tag{3.27}$$

Define $l_j = \max_{0 \leq l \leq r} \{l : t(k_j + l) = p\}$. The existence of l_j is guaranteed by almost cyclicality of the control sequence. Let $s_j = k_j + l_j$ and assume that $c_i^{s_j} = \theta_i^{s_j}$. Then

$$\langle a^i, w_i^{s_j} \rangle = \alpha^{s_j} b_i + (1 - \alpha^{s_j}) \langle a^i, x^{s_j} \rangle$$

i.e.,

$$\langle a^i, w_i^{s_j} - x^{s_j} \rangle = \alpha^{s_j} \left(b_i - \langle a^i, x^{s_j} \rangle \right).$$

Therefore

$$\alpha_i^{s_j} \left(b_i - \langle a^i, x^* \rangle \right) = \langle a^i, w_i^{s_j} - x^{s_j} \rangle + \alpha^{s_j} \langle a^i, x^{s_j} - x^* \rangle.$$

Thus, using (3.27) and the fact that $0 < \epsilon \leq \alpha^{s_j} \leq 1$, we have the following inequalities for all $j \geq J$ and $i \in I_1^{t(s_j)}(x^*)$:

$$\begin{aligned}
\epsilon \left(b_i - \langle a^i, x^* \rangle \right) &\leq \langle a^i, w_i^{s_j} - x^{s_j} \rangle + \alpha^{s_j} \langle a^i, x^{s_j} - x^* \rangle \\
&\leq \|a^i\| \left(\|w_i^{s_j} - x^{s_j}\| + \alpha^{s_j} \|x^{s_j} - x^*\| \right) \\
&\leq \|a^i\| \left(\|w_i^{s_j} - x^*\| + \|x^{s_j} - x^*\| + \alpha^{s_j} \|x^{s_j} - x^*\| \right) \\
&< \|a^i\|(2 + \alpha^{s_j})\delta \leq 3\delta \|a^i\|.
\end{aligned}$$

Hence we have the contradiction $3\delta > \epsilon \{(b_i - \langle a^i, x^* \rangle)/\|a^i\|\} \geq 4\delta$ for all $i \in I_1^{t(s_j)}(x^*)$. It therefore follows that $c_i^{s_j} \neq \theta_i^{s_j}$ and so $c_i^{s_j} = \frac{z_i^{s_j}}{\lambda_i^{s_j}}$ implies $z_i^{s_j+1} = 0$ for $i \in I_1^{t(s_j)}(x^*) = I_1^p(x^*)$. By the definition of l_j, the index p is not used in iteration $k_j + l$ for $l_j < l \leq r$ and so $z_i^{s_j+l}$ for $i \in I_1^p(x^*)$ remains unaffected. We conclude that $z_i^{s_j+r+1} = 0$ for $i \in I_1^p(x^*)$ with $p \in \{1, \ldots, M\}$. □

Now the proof of Theorem 3.2.1 follows:

Proof. Since $z_i^{k_j} = 0$ for $i \in I_1^{t(k_j)}(x^*)$ whiles $\langle a^i, x^* \rangle = b_i$ for $i \in I_2^{t(k_j)}(x^*)$, and $I_{t(k_j)} = I_1^{t(k_j)} \cup I_2^{t(k_j)}$ for all $j \geq 0$, we have

$$\begin{aligned} \langle z^{k_j}, Ax^{k_j} - b \rangle &= \langle z^{k_j}, Ax^{k_j} - Ax^* \rangle \\ &= \langle A^T z^{k_j}, x^{k_j} - x^* \rangle \end{aligned}$$

But $A^T z^{k_j} = -\nabla f(x^{k_j})$ and so

$$\begin{aligned} \langle z^{k_j}, Ax^{k_j} - b \rangle &= -\langle \nabla f(x^{k_j}), x^{k_j} - x^* \rangle \\ &= D_f(x^*, x^{k_j}) - f(x^*) + f(x^{k_j}) \to 0 \text{ as } j \to \infty. \end{aligned}$$

Therefore by the continuity of f on $\bar{S}$, we have

$$\lim_{j\to\infty} L(x^{k_j}, z^{k_j}) = \lim_{j\to\infty} \left(f(x^{k_j}) + \langle z^{k_j}, Ax^{k_j} - b \rangle \right) = f(x^*).$$

But $L(x^{k_j}, z^{k_j}) \leq f(z) - D_f(z, x^{k_j}) \leq f(z)$ for all $z \in C \cap \bar{S}$. Therefore $f(x^*) \leq f(z)$ for any $z \in C \cap \bar{S}$. This shows that x^* is a solution for (1.43)-(1.45), since by Proposition 3.2.6, $x^* \in C$. This solution is unique because of the strict convexity of f. □

Chapter 4

Closed form formulas for separated variables optimization

From the computational point of view, the difficult part of any Bregman's algorithm lies in the projection operation at each iterative step. For n-dimensional iterates, the system of equations to solve at each iterative step usually consists of $n+1$ equations, n of which are usually nonlinear. This means that if the system has to be solved numerically in each iteration, then the computational burden might reduce the efficiency of the algorithm. Numerical errors in the calculation of the projections may also cause the practical algorithm to deviate from the conceptual one.

To reduce the computational burden, we develop a closed-form formula for the iterative step in Bregman's algorithm for the optimization of any Bregman function over linear constraints. That is, we replace the computational burden involved in an inner loop calculation of the projection parameter by a closed-form formula. In [23] and [22], closed-form formulas were derived for the iterative steps for the maximization of Burg's and Shannon's entropy. It was also established in [23] that MART step is indeed a secant approximation to Bregman's iterative step for entropy maximization and this is the motivation for the work in this chapter.

4.1 Analysis of Bregman's algorithm for optimization of variable separable functions

We consider the minimization of the Bregman function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = \sum_{j=1}^{n} g_j(x_j)$ with zone $S = \text{Int}(\text{dom}f)$. We assume that $g_j'' > 0$ everywhere and that f is essentially smooth, twice continuously differentiable and zone consistent with respect to the hyperplane $H_i = \{x \in \mathbb{R}^n \mid \langle a^i, x\rangle = b_i\}$ for $i = 1, \ldots, m$ and $\cap_{i=1}^m H_i \neq \emptyset$.

We begin with the problem

$$\min f(x) \text{ subject to } Ax = b, \tag{4.1}$$

where A is an $m \times n$ real matrix; x and b are n and m dimensional vectors respectively. $a^i \neq 0$ for $i = 1, \ldots, m$ is the transpose of the ith row of A and b_i the ith component of b.

If x^k is the current iterate then, according to Lemma 1.6.10, the following equations determine uniquely the next iterate x^{k+1} and the projection parameter c_k in the Bregman's algorithm for solving (4.1) at kth iterative step.

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

This implies

$$g_j'(x_j^{k+1}) = g_j'(x_j^k) + c_k a_j^i, \tag{4.2}$$

$$\langle a^i, x^{k+1}\rangle = b_i. \tag{4.3}$$

We assume that the function $f \in B(S)$ satisfies Assumption 2.2.2 since the applicability of the algorithm defined by (4.2) and (4.3) depends on the ability to invert the gradient ∇f explicitly.

Now eliminating x^{k+1} from these two equations, we have

$$\sum_{j=1}^{n} a_j^i G_j^{-1}\left(G_j(x_j^k) + c a_j^i\right) = b_i, \text{ where } c_k = c, \quad g_j' = G_j \tag{4.4}$$

and the inverse G_j^{-1} exists based on the assumption that f is essentially smooth and zone consistent with respect to the hyperplane $H_i = \{x \in \mathbb{R}^n \mid \langle a^i, x\rangle = b_i\}$.

We define the function $h : \mathbb{R} \to \mathbb{R}$ by

$$h(c) = \sum_{j=1}^{n} a_j^i G_j^{-1}\left(G_j(x_j^k) + c a_j^i\right) - b_i. \tag{4.5}$$

We find the approximate root of (4.5) using the root of a secant line to $h(c)$.
To verify whether $h(c)$ really has a root, we consider the following features:

$$h'(c) = \sum_{j=1}^{n}(a_j^i)^2 (G_j^{-1})'\left(G_j(x_j^k) + c a_j^i\right)$$

and

$$(G_j^{-1})'\left(G_j(x_j^k) + c a_j^i\right) = \frac{1}{G_j'\left(G_j^{-1}\left(G_j(x_j^k) + c a_j^i\right)\right)} = \frac{1}{g''\left(G_j^{-1}(G_j\left(x_j^k\right) + c a_j^i)\right)}$$

which is positive since g_j is a strictly convex and twice continuously differentiable function. Therefore $h'(c) > 0$ for all $c \in \mathbb{R}$. This means that h is monotonically increasing on its domain, and by Lemma 1.6.10, the system (4.2)-(4.3) also determines uniquely the parameter c. Thus $h(c)$ has unique real root.

Interestingly, these features of $h(c)$ remain unchanged with relaxation, i.e., if we substitute b_i in (4.5) for $\alpha_i^k b_i + (1 - \alpha_i^k)\langle a^i, x^k\rangle$, where the relaxation parameter α_i^k is such that $0 < \alpha_i^k < 1$ for all $k \geq 0$ and $i = 1, \dots, m$.

To find an approximate value for the root of $h(c) = 0$, we first find a secant line to the curve $h(c)$ as follows.
The linear approximation of $G_j^{-1}\left(G_j(x_j^k) + c a_j^i\right)$ is given by

$$G_j^{-1}\left(G_j(x_j^k) + c a_j^i\right) \approx G_j^{-1}(G_j(x_j^k)) + c a_j^i (G_j^{-1})'(G_j(x_j^k)) = x_j^k + c\frac{a_j^i}{g_j''(x_j^k)}. \tag{4.6}$$

We substitute (4.6) into (4.5) and the resulting expression becomes the secant line $\hat{f}(c)$ to the curve $h(c)$. That is

$$\hat{f}(c) = \sum_{j=1}^{n} a_j^i \left(x_j^k + c\frac{a_j^i}{g_j''(x_j^k)}\right) - b_i = \langle a^i, x^k\rangle - b_i + c\sum_{j=1}^{n}\frac{(a_j^i)^2}{g_j''(x_j^k)}$$

and for the c-intercept, $\bar{c}$, we have

$$\bar{c} = \frac{b_i - \langle a^i, x^k\rangle}{\sum_{j=1}^{n}\frac{(a_j^i)^2}{g_j''(x_j^k)}}$$

and with underrelaxation (i.e., replacing b_i with $\alpha_i^k b_i + (1-\alpha_i^k)\langle a^i, x^k\rangle$), we have

$$\bar{c} = \alpha_i^k \frac{b_i - \langle a^i, x^k\rangle}{\sum_{j=1}^n \frac{(a_j^i)^2}{g_j''(x_j^k)}}. \tag{4.7}$$

Therefore using the last equation and (4.2), the closed-form formula in the iterative step in underrelaxed Bregman's algorithm for solving (4.1) is

$$x_j^{k+1} = G_j^{-1}\left(g_j'(x_j^k) + a_j^i \alpha_i^k \frac{b_i - \langle a^i, x^k\rangle}{\sum_{j=1}^n \frac{(a_j^i)^2}{g_j''(x_j^k)}}\right) \tag{4.8}$$

for $k \geq 0$, $j = 1, \ldots, n$ and the control sequence $\{i(k)\}$ is almost cyclic on $\{1, 2, \ldots, m\}$.

We therefore propose the following Bregman's algorithm which employs closed-form formula for the iterative updates for solving (4.1).

4.1.1 Bregman's algorithm for linear equalities using closed-form formula

Algorithm 4.1.1. Bregman's algorithm for linear equalities using closed-form formula

(i) **Initialization** $x^0 \in \text{Intdom} f$ is such that for an arbitrary $z^0 \in \mathbb{R}_+^m$,

$$\nabla f(x^0) = -A^T z^0.$$

(ii) **Iterative Step**

$$x_j^{k+1} = G_j^{-1}\left(g_j'(x_j^k) + a_j^i \alpha_i^k \frac{b_i - \langle a^i, x^k\rangle}{\sum_{j=1}^n \frac{(a_j^i)^2}{g_j''(x_j^k)}}\right)$$

The sequence $\{i(k)\}$ is almost cyclic on the index set $\{1, 2, \ldots, m\}$ and $\{\alpha_i^k\}$ is a sequence of relaxation parameters such that $\epsilon \leq \alpha_i^k \leq 1$ for a fixed $\epsilon > 0$.

$\diamondsuit$

4.1.2 General underrelaxed Bregman's algorithm for linear inequalities

The algorithm for linear inequality constraints also calculates the projection parameter $\bar{c}$ in (4.7). However, before proceeding, it compares it with the ith component of the current dual vector z^k and uses the smaller of the two in the iteration. We therefore propose the following general Bregman algorithm for the minimization of any Bregman function over inequality constraints with the projection parameter given in a closed-form in (4.7). That is, a general underrelaxed Bregman's algorithm for solving the problem

$$\min f(x) \text{ subject to } Ax \leq b,\ x \in \bar{S} \tag{4.9}$$

where A and b are as defined in (4.1).

Algorithm 4.1.2. General underrelaxed Bregman's algorithm for linear inequalities

(i) **Initialization** $x^0 \in \text{Intdom} f$ is such that for an arbitrary $z^0 \in \mathbb{R}^m_+$,

$$\nabla f(x^0) = -A^T z^0.$$

(ii) **Iterative Step** Given x^k and z^k, calculate x^{k+1} and z^{k+1} from

$$\begin{aligned} \nabla f(x^{k+1}) &= \nabla f(x^k) + c_i^k a^i, \\ z^{k+1} &= z^k - c_i^k e^i \end{aligned}$$

with

$$c_i^k = \min\left(z_i^k, \theta_i^k\right)$$

where

$$\theta_i^k = \alpha_i^k \frac{b_i - \langle a^i, x^k\rangle}{\sum_{j=1}^n (a_j^i)^2 / g_j''(x_j^k)}.$$

(iii) $H(k)_i = \{x \in \mathbb{R}^n \mid \langle a^i, x\rangle = \alpha_i^k b_i + (1-\alpha_i^k)\langle a^i, x^k\rangle\}$.

(iv) The sequence $\{i(k)\}$ is almost cyclic on the index set $\{1, 2, \ldots, m\}$

and $\{\alpha_i^k\}$ is a sequence of relaxation parameters such that $\epsilon \leq \alpha_i^k \leq 1$ for a fixed $\epsilon > 0$. We assume that the problem is feasible and f is strongly zone consistent with respect to every H_i. $\diamondsuit$

Note: By analyzing Lemma 4 and its proof in [23] and the proof of Theorem 1 which is the main result of [82], we conjecture that algorithms 4.1.1 and 4.1.2 converge to the desired solutions of problems (4.1) and (4.9).

We will now use (4.8) to derive estimates for closed-form formulas for the iterative steps in Bregman's algorithm for the following functions:

4.1.3 The half-squared Euclidean norm

As noted in Chapter 1, the function $f = \frac{1}{2}\|\cdot\|^2$, which leads to orthogonal projections onto hyperplanes, always leads to a closed-form formula for the iterative step. That is the Kaczmarz's algorithm for solving the system $Ax = b$. When $g_j(x_j^k) = \frac{1}{2}(x_j^k)^2$ in (4.8), we recover exactly the closed-form formula for Kaczmarz's algorithm.

Thus, for $g_j(x_j^k) = \frac{1}{2}(x_j^k)^2$, we have

$$g_j'(x_j^k) = x_j^k \ , \ g_j''(x_j^k) = 1 \text{ and } G_j^{-1}(x_j^k) = x_j^k \tag{4.10}$$

and the substitution of (4.10) into (4.8) gives

$$x_j^{k+1} = x_j^k + a_j^i \alpha_i^k \frac{b_i - \langle a^i, x^k \rangle}{\sum_{j=1}^n (a_j^i)^2} = x_j^k + \alpha_i^k \frac{b_i - \langle a^i, x^k \rangle}{\|a^i\|^2} a_j^i$$

or

$$x^{k+1} = x^k + \alpha_i^k \frac{b_i - \langle a^i, x^k \rangle}{\|a^i\|^2} a^i$$

for all $k \geq 0$.

4.1.4 The negative Shannon entropy

We use (4.8) to derive a closed-form formula for the case when $g_j(x_j^k) = x_j^k \ln x_j^k$ for $x_j^k > 0$ for each j and establish its relation with MART. Now, for $g_j(x_j^k) = x_j^k \ln x_j^k$, we have

$$g_j'(x_j^k) = 1 + \ln x_j^k \ , \ g_j''(x_j^k) = \frac{1}{x_j^k} \text{ and } G_j^{-1}(x_j^k) = \exp(x_j^k - 1) \tag{4.11}$$

and the substitution of (4.11) into (4.8) gives

$$\begin{aligned} x_j^{k+1} &= G_j^{-1}\left(1 + \ln x_j^k + a_j^i \alpha_i^k \frac{b_i - \langle a^i, x^k \rangle}{\sum_{j=1}^n (a_j^i)^2 x_j^k}\right) \\ &= \exp\left(\ln x_j^k + a_j^i \alpha_i^k \frac{b_i - \langle a^i, x^k \rangle}{\sum_{j=1}^n (a_j^i)^2 x_j^k}\right) \\ &= x_j^k \exp\left(a_j^i \alpha_i^k \frac{b_i - \langle a^i, x^k \rangle}{\sum_{j=1}^n (a_j^i)^2 x_j^k}\right) \\ &= x_j^k \exp \frac{a_j^i \alpha_i^k \langle a^i, x^k \rangle}{\sum_{j=1}^n (a_j^i)^2 x_j^k}\left(\frac{b_i}{\langle a^i, x^k \rangle} - 1\right). \end{aligned} \tag{4.12}$$

For the convergence of MART or Bregman's algorithm for entropy maximization over equality constraints, the following assumptions are made:

(i) $a_j^i > 0$ for $i = 1, \ldots, m$, and $j = 1, \ldots, n$.

(ii) $Ax = b$ is scaled so that $a_j^i \leq 1$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

Using these assumptions, $0 < a_j^i \leq 1$ implies $0 < (a_j^i)^2 x_j^k \leq a_j^i x_j^k$ and so

$$0 \leq \sum_{j=1}^n (a_j^i)^2 x_j^k \leq \langle a^i, x^k \rangle \text{ or } \frac{1}{\sum_{j=1}^n (a_j^i)^2 x_j^k} \geq \frac{1}{\langle a^i, x^k \rangle}. \tag{4.13}$$

Therefore, for the case $\frac{b_i}{\langle a^i, x^k \rangle} > 1$, multiplying the last inequality by $\alpha_i^k(b_i - \langle a^i, x^k \rangle)$, we have

$$a_j^i \alpha_i^k \frac{b_i - \langle a^i, x^k \rangle}{\sum_{j=1}^n (a_j^i)^2 x_j^k} \geq a_j^i \alpha_i^k \frac{b_i - \langle a^i, x^k \rangle}{\langle a^i, x^k \rangle} = a_j^i \alpha_i^k \left(\frac{b_i}{\langle a^i, x^k \rangle} - 1\right).$$

Now, using the inequality $\ln y \leq y - 1$ for $y > 1$, we have

$$a_j^i \alpha_i^k \left(\frac{b_i}{\langle a^i, x^k \rangle} - 1\right) \geq a_j^i \alpha_i^k \ln\left(\frac{b_i}{\langle a^i, x^k \rangle}\right) = \ln\left(\frac{b_i}{\langle a^i, x^k \rangle}\right)^{a_j^i \alpha_i^k}.$$

Using the last inequality and (4.12), we have

$$\begin{aligned} x_j^{k+1} &\geq \exp\left(\ln x_j^k + \ln\left(\frac{b_i}{\langle a^i, x^k \rangle}\right)^{a_j^i \alpha_i^k}\right) \\ &= x_j^k \left(\frac{b_i}{\langle a^i, x^k \rangle}\right)^{a_j^i \alpha_i^k}. \end{aligned}$$

This means that for the case $\frac{b_i}{\langle a^i, x^k \rangle} > 1$, our formula produces iterates that majorize those generated by MART.

Now for the case $0 < \frac{b_i}{\langle a^i, x^k \rangle} < 1$, multiplying (4.13) by $\alpha_i^k(b_i - \langle a^i, x^k \rangle)$, we have

$$a_j^i \alpha_i^k \frac{b_i - \langle a^i, x^k \rangle}{\sum_{j=1}^n (a_j^i)^2 x_j^k} \leq a_j^i \alpha_i^k \frac{b_i - \langle a^i, x^k \rangle}{\langle a^i, x^k \rangle} = \frac{a_j^i b_i \alpha_i^k}{\langle a^i, x^k \rangle} \left(1 - \frac{\langle a^i, x^k \rangle}{b_i} \right)$$

and using the inequality $1 - \frac{1}{y} \leq \ln y$ for $0 < y \leq 1$, we have

$$a_j^i \alpha_i^k \frac{b_i - \langle a^i, x^k \rangle}{\sum_{j=1}^n (a_j^i)^2 x_j^k} \leq \frac{a_j^i b_i \alpha_i^k}{\langle a^i, x^k \rangle} \ln \left(\frac{b_i}{\langle a^i, x^k \rangle} \right) = \alpha_i^k \frac{b_i}{\langle a^i, x^k \rangle} \ln \left(\frac{b_i}{\langle a^i, x^k \rangle} \right)^{a_j^i}.$$

If we define a sequence of relaxation $\{\lambda_i^k\}$ such that $\lambda_i^k = \alpha_i^k \frac{b_i}{\langle a^i, x^k \rangle}$ then (4.12) becomes

$$\begin{aligned} x_j^{k+1} \quad &\leq \quad \exp \left(\ln x_j^k + \lambda_i^k \ln \left(\frac{b_i}{\langle a^i, x^k \rangle} \right)^{a_j^i} \right) \\ &= x_j^k \left(\frac{b_i}{\langle a^i, x^k \rangle} \right)^{a_j^i \lambda_i^k}. \end{aligned}$$

This means that for the case $0 < \frac{b_i}{\langle a^i, x^k \rangle} < 1$, our formula produces iterates that are majorized by those of MART.

However, in order to compare the efficiency of the two algorithms, we ought to do some numerical experimentation on both MART and Algorithm 4.1.2.

4.1.5 The negative Burg's entropy

The Burg's entropy, $B(x)$, also known as '$\log x$-entropy' is defined on $\mathbb{R}_{++}^n$ by

$$B(x) = \sum_{j=1}^n \log x_j.$$

Burg's entropy was first proposed in [17] and has since then provoked a controversy regarding the question of which entropy functional should be used in different situations. This question was discussed in [46] and recently also in [66]. It must be noted that the negative Burg's entropy is not a Bregman function because it becomes singular on the boundary of its zone, i.e., when x_j^k tends to zero for even only one j, then $B(x)$ tends to ∞, demonstrating an essential discontinuity.

In this subsection, we use (4.8) to derive a closed-form formula for the minimization of the negative Burg's entropy, i.e., $f(x) = -B(x)$ with zone $S = \mathrm{Int}\mathbb{R}^n_+$ and $g_j(x^k_j) = -\ln x^k_j$ for $x^k_j > 0$ and for each j. We will compare our method with the method in [22].

Now for $g_j(x^k_j) = -\ln x^k_j$, we have

$$g'_j(x^k_j) = -\frac{1}{x^k_j}\ ,\ g''_j(x^k_j) = \frac{1}{(x^k_j)^2} \text{ and } G^{-1}_j(x^k_j) = -\frac{1}{x^k_j}. \tag{4.14}$$

The substitution of (4.14) into (4.8) gives

$$\begin{aligned} x^{k+1}_j &= G^{-1}_j\left(-\frac{1}{x^k_j} + a^i_j\alpha^k_i\frac{b_i - \langle a^i, x^k\rangle}{\sum_{j=1}^n (a^i_j x^k_j)^2}\right) \\ &= \frac{x^k_j\sum_{j=1}^n (a^i_j x^k_j)^2}{\sum_{j=1}^n (a^i_j x^k_j)^2 + (a^i_j\alpha^k_i x^k_j)(\langle a^i, x^k\rangle - b_i)} \\ &= \frac{x^k_j}{1-u^k_j}, \text{ where } u^k_j = \frac{a^i_j\alpha^k_i x^k_j(b_i - \langle a^i, x^k\rangle)}{\sum_{j=1}^n (a^i_j x^k_j)^2} \neq 1. \end{aligned} \tag{4.15}$$

In [22], using underrelaxation and with the condition $a^i_j b_i \geq 0$ imposed on the elements of the matrix A and the vector b, the following expression was derived for the projection parameter c_k.

$$\varphi(a^i, x^k, b_i) = \begin{cases} \lambda_k\left(1 - \frac{\langle a^i, x^k\rangle}{b_i}\right) t_k \text{ if } b_i > 0, \\ \lambda_k\left(1 - \frac{\langle a^i, x^k\rangle}{b_i}\right) r_k \text{ if } b_i < 0, \end{cases} \tag{4.16}$$

where $0 < \lambda_k < 1$, and

$$r_k = \max\left\{\frac{1}{a^i_j x^k_j} \mid 1 \leq j \leq n,\ a^i_j < 0\right\}$$

and if $a^i_j \geq 0$ for all j then $r_k = -\infty$;

$$t_k = \min\left\{\frac{1}{a^i_j x^k_j} \mid 1 \leq j \leq n,\ a^i_j > 0\right\}$$

and if $a^i_j \leq 0$ for all j then $t_k = \infty$.

But our projection parameter c_k with relaxation is estimated here as

$$c_k = \alpha^k_i\frac{b_i - \langle a^i, x^k\rangle}{\sum_{j=1}^n (a^i_j x^k_j)^2} = \frac{b_i\alpha^k_i}{\sum_{j=1}^n (a^i_j x^k_j)^2}\left(1 - \frac{\langle a^i, x^k\rangle}{b_i}\right). \tag{4.17}$$

We compare (4.17) with (4.16) in [22] and derive any relation between the two estimates.

By the definition of t_k in [22], $a^i_j x^k_j \leq \frac{1}{t_k}$ if $a^i > 0$, and $t_k = \infty$ if $a^i_j \leq 0$ for all j. This means that, $(a^i_j x^k_j)^2 \leq \frac{a^i_j x^k_j}{t_k}$ or $\sum_{j=1}^n (a^i_j x^k_j)^2 \leq \frac{\langle a^i, x^k \rangle}{t_k}$.

Therefore
$$\frac{b_i}{\sum_{j=1}^n (a^i_j x^k_j)^2} \geq \frac{t_k b_i}{\langle a^i, x^k \rangle} \text{ for } b_i > 0 \text{ and } a^i_j \neq 0$$
since $a^i_j b_i > 0$. Thus, using (4.17) for $\frac{b_i}{\langle a^i, x^k \rangle} < 1$, we have

$$c_k \leq \frac{b_i}{\langle a^i, x^k \rangle} \alpha^k_i \left(1 - \frac{\langle a^i, x^k \rangle}{b_i}\right) t_k \tag{4.18}$$

and for $\frac{b_i}{\langle a^i, x^k \rangle} > 1$, we have

$$c_k \geq \frac{b_i}{\langle a^i, x^k \rangle} \alpha^k_i \left(1 - \frac{\langle a^i, x^k \rangle}{b_i}\right) t_k. \tag{4.19}$$

Similarly, by the definition of r_k in [22], $a^i_j x^k_j \geq \frac{1}{r_k}$ if $a^i < 0$, and $r_k = -\infty$ if $a^i_j \geq 0$ for all j. This means that, $(a^i_j x^k_j)^2 \leq \frac{a^i_j x^k_j}{r_k}$ or $\sum_{j=1}^n (a^i_j x^k_j)^2 \leq \frac{\langle a^i, x^k \rangle}{r_k}$.

Therefore
$$\frac{b_i}{\sum_{j=1}^n (a^i_j x^k_j)^2} \leq \frac{r_k b_i}{\langle a^i, x^k \rangle} \text{ for } b_i < 0$$
since $a^i_j b_i \geq 0$. Thus, using (4.17) for $\frac{b_i}{\langle a^i, x^k \rangle} < 1$, we have

$$c_k \geq \frac{b_i}{\langle a^i, x^k \rangle} \alpha^k_i \left(1 - \frac{\langle a^i, x^k \rangle}{b_i}\right) r_k \tag{4.20}$$

and for $\frac{b_i}{\langle a^i, x^k \rangle} > 1$, we have

$$c_k \leq \frac{b_i}{\langle a^i, x^k \rangle} \alpha^k_i \left(1 - \frac{\langle a^i, x^k \rangle}{b_i}\right) r_k. \tag{4.21}$$

Chapter 5

Analysis of inconsistent problems

5.1 Introduction

As stated in the introduction, ART is a very well known iterative algorithm for image processing and reconstruction [51] that aims at solving a linear system of equations

$$Ax = b, \tag{5.1}$$

where $A = (a^i_j)$ is the $m \times n$ nonnegative projection matrix, x is the image vector with components x_j and b is the m-dimensional vector of projection data with ith element b_i $(b_i \geq 0)$. For a given starting point x^0, ART's iteration is defined by

$$x^{k+1} = x^k + \lambda_k \frac{b_i - \langle a^i, x^k \rangle}{\|a^i\|^2} a^i. \tag{5.2}$$

$a^i \neq 0$ is the ith column of the transpose A^T and λ_k is a positive relaxation parameter that lies in the open interval $(0, 2)$. The sequence $\{i(k)\}$ determines the ordering (usually cyclic) in which the matrix rows are selected [21].

In the consistent case, when the system (5.1) has solutions, (5.2) converges to the solution for which $\|x - x^0\|$ is minimized. If the system (5.1) has no solution, ordering is cyclic and the relaxation parameter is fixed, it has been proven [25] that (5.2) generates m fixed points, that, when λ tends to zero, (5.2) approaches a weighted least squares solution, that is, the limit solves the problem

$$\min_{x \geq 0} \sum_{i=1}^{m} \left(\frac{b_i - \langle a^i, x \rangle}{\|a^i\|} \right)^2. \tag{5.3}$$

In [24, 81], it was also proven that the sequence (5.2) itself converges to the solution of (5.3) provided that the relaxation parameters satisfy the conditions

$$\lambda_k \underset{k\to\infty}{\longrightarrow} 0 \tag{5.4}$$

and

$$\sum_{k=0}^{\infty} \lambda_k = +\infty. \tag{5.5}$$

Block versions of ART share the same properties mentioned above and the fully parallel version (a single block), known as Cimmino's method can be shown to converge to the solution in the feasible case and to a least squares solution in the infeasible case for a fixed relaxation parameter in (0, 2) (see [40]).

The question that follows is that, is it true that for other Bregman methods or algorithms, the under-relaxed version of the sequences generated by the algorithms in the inconsistent case converges to the solution of an optimization problem, as it is the case in ART, if the relaxation parameters satisfy certain condition? Also can the fully simultaneous version of the sequences generated by other Bregman methods in the inconsistent case converge to the solution of an optimization problem? In this chapter, we give an answer for the particular case of MART and its simultaneous version SMART, already analyzed in [18].

One of the first iterative algorithms in image reconstruction for underdetermined problems (limited number of views) is MART [51]. This is also a very well known nonlinear iterative algorithm for transmission computed tomography (CT) with very attractive properties. It converges to a maximum entropy solution of the linear CT equations [35, 71, 37] and it confines the reconstruction to the convex hull of the object [8, 33].

Both ART and MART were introduced by Gordon, Bender and Herman. But MART is limited to non-negative systems for which non-negative solutions are sought. That is, MART finds non-negative solutions to the system (5.1) provided (5.1) is non-negative, i.e., a^i_j's are non-negative and b_i's are positive.

In the under-determined case, both algorithms find the solution closest to the starting point, in the two-norm or weighted two-norm sense for ART, and in the cross-entropy sense for MART. Thus, both algorithms can be viewed as solving optimization problem.

MART (now we consider the cyclic version, $i(k) = k(\text{mod } m)$) is defined by the following sequence: given a positive starting point x^0, and, in general, $x^k = x^{(k,0)}$, and $x^{k+1} = x^{(k,m)}$, then, for all $i = 1, \ldots, m$ and, $j = 1, \ldots, n$ such that $a^i_j > 0$ and $x^k_j > 0$,

$$x_j^{(k,i)} = x_j^{(k,i-1)} \left(\frac{b_i}{\langle a^i, x^{(k,i-1)} \rangle} \right)^{\lambda_k a^i_j}. \tag{5.6}$$

Like ART, in the consistent case, when the system (5.1) has solutions, (5.6) converges to the solution for which $K(x, x^0)$ is minimized, where the function K is as defined in (5.8) below.

Next, we prove that the properties of MART are similar to those of ART when the problem is inconsistent. We prove that, when properly underrelaxed as in (5.4) and (5.5), the algorithm converges to a solution of the problem

$$\min_{x \geq 0} L(x) := \sum_{i=1}^{m} \left(\langle a^i, x \rangle \ln \langle a^i, x \rangle - \langle a^i, x \rangle (\ln b_i + 1) \right) \tag{5.7}$$

which is the projection of the data vector b onto the range of A with respect to the Kullback-Leibler distance [69]. The Kullback-Leibler distance, or relative entropy of the vector x in $\mathbb{R}^n_+$ with respect to the vector y also in $\mathbb{R}^n_{++}$ is defined as

$$K(x, y) = \sum_{j=1}^{n} \left(x_j \ln \frac{x_j}{y_j} + y_j - x_j \right). \tag{5.8}$$

So, (5.7) is equivalent to

$$\min_{x \geq 0} K(Ax, b), \tag{5.9}$$

since, using (5.8) with the assumption that $a^i_j \geq 0$ and $b_i > 0$, we have $Ax \in \mathbb{R}^m_+$ and $b \in \mathbb{R}^m_{++}$ and so

$$K(Ax, b) = \sum_{i=1}^{m} (\langle a^i, x \rangle \ln \langle a^i, x \rangle - \langle a^i, x \rangle (\ln b_i + 1) + b_i), \tag{5.10}$$

which has the same minimizer as (5.7) since b_i is a positive constant for each i. The solution of (5.9) is not necessarily unique. We will prove that MART chooses the one minimizing $K(x, x^0)$, where x^0 is the (positive) starting point. It gives the maximum entropy if $x^0 = \mathbf{1}$. We will confine our results to row action MART described above, for the sake of simplicity. Proofs for more general block versions, like those in [18], are similar.

5.2 Convergence results

5.2.1 Boundedness

Taking into account (5.4) for k large enough, we have that $\forall\ i, j$

$$\lambda_k a_j^i \ \leq \ 1. \tag{5.11}$$

Define

$$c \ = \ \min_{a_j^i \neq 0}\{a_j^i\} \tag{5.12}$$

and

$$\bar{b} \ = \ \max_i \, b_i. \tag{5.13}$$

Then, using (5.11) and the fact that $x_j^k > 0$ for each j and $k \geq 0$, we obtain the following inequalities for $a_j^i > 0$ (when $a_j^i = 0$ the iteration remains unmodified) and $\frac{b_i}{\langle a^i, x^k\rangle} > 1$.

$$\begin{aligned} 0 \ < \ x_j^{k+1} &= x_j^k \left(\frac{b_i}{\langle a^i, x^k\rangle}\right)^{\lambda_k a_j^i} \leq x_j^k \frac{b_i}{\langle a^i, x^k\rangle} \\ &\leq \frac{x_j^k a_j^i b_i}{c\sum_l a_l^i x_l^k} \leq \frac{b_i}{c\left(1 + \sum_{l\neq j} \frac{a_l^i x_l^k}{a_j^i x_j^k}\right)} \leq \frac{\bar{b}}{c}. \end{aligned} \tag{5.14}$$

The inequalities in (5.14) are as a result of (5.12) and (5.13).

Now suppose that, for some k, $x_j^k > \frac{\bar{b}}{c}$, then the inequality (5.14) is valid if $\frac{b_i}{\langle a^i, x^k\rangle} > 1$, and so $\frac{\bar{b}}{c} < x_j^k < x_j^k \frac{b_i}{\langle a^i, x^k\rangle} \leq \frac{\bar{b}}{c}$ which is a contradiction. Hence $x_j^k \leq \frac{\bar{b}}{c}$. Also if $\frac{b_i}{\langle a^i, x^k\rangle} \leq 1$, then from (5.6), $x_j^{k+1} \ \leq \ x_j^k$. Using induction on k, it can be easily proved that $x_j^k \ \leq \ x_j^{k_0}$ for some given k_0. So, the sequence generated by (5.6) is uniformly bounded (independently of λ).

5.2.2 Change of variables

Now, let us change variables in (5.6). Taking logarithms and using the notation $e^y = (e^{y_1}, \dots, e^{y_n})^T$ for $y = (y_1, \dots, y_n)^T$ and $y_j^{(k,i)} = \ln\left(x_j^{(k,i)}\right)$, we obtain the iteration

$$y_j^{(k,i)} = y_j^{(k,i-1)} + \lambda_k a_j^i \ln\left(\frac{b_i}{\langle a^i, e^{y^{(k,i-1)}}\rangle}\right) \text{ for } j = 1, \dots, n. \tag{5.15}$$

Alternatively, we may consider $K(x_j^{(k,i)}, 1)$ and use $\nabla K(x_j^{(k,i)}, 1)$ for $y_j^{(k,i)}$ since $K(x_j^{(k,i)}, 1) = x_j^{(k,i)} \ln x_j^{(k,i)} + 1 - x_j^{(k,i)}$ and $\nabla K(x_j^{(k,i)}, 1) = \ln x_j^{(k,i)} = y_j^{(k,i)}$. (5.15) therefore takes the form

$$\nabla K(x_j^{(k,i)}, 1) = \nabla K(x_j^{(k,i-1)}, 1) + \lambda_k a_j^i \ln\left(\frac{b_i}{\langle a^i, x^{(k,i-1)}\rangle}\right) \text{ for } j = 1, \dots, n. \tag{5.16}$$

Now, summing (5.15) over i for $j = 1, \dots, n$, we have

$$y_j^{(k,m)} - y_j^{(k,0)} = \lambda_k \sum_{i=1}^{m} a_j^i \ln\left(\frac{b_i}{\langle a^i, e^{y^{(k,i-1)}}\rangle}\right). \tag{5.17}$$

(5.15) and (5.17) motivate the definition of the following objective function

$$\hat{L}(y) := \sum_{i=1}^{m} (\langle a^i, e^y\rangle \ln\langle a^i, e^y\rangle - \langle a^i, e^y\rangle(\ln b_i + 1)). \tag{5.18}$$

Equivalently, we write

$$\hat{L}(y) = \sum_{i=1}^{m} \hat{L}_i(y), \tag{5.19}$$

where

$$\hat{L}_i(y) = \langle a^i, e^y\rangle \ln\langle a^i, e^y\rangle - \langle a^i, e^y\rangle(\ln b_i + 1). \tag{5.20}$$

In order to simplify our notation, we will consider the extended n-dimensional vector space, adding possibly $-\infty$ coordinates, that we will denote by $\mathcal{R}_-^n$. There is a one-to-one correspondence between this set and $\mathbb{R}_+^n$ via the logarithm of the coordinates.

A trivial observation using (5.7) and (5.19) is that,

$$\min_{x \geq 0} L(x) = \min_{y \in \mathbb{R}^n} \hat{L}(y) \tag{5.21}$$

and for $x = e^y$,

$$e^{-y_j} \nabla \hat{L}(y)_j = \nabla L(x)_j. \tag{5.22}$$

It is easy to show that, if $y^* \in \mathcal{R}^n_-$ is a minimum, and $x^* = e^{y^*}$, then, for $i = 1, \ldots, m$, using (5.22),

$$\nabla \hat{L}(y^*)_j \; = \; x^*_j \nabla L(x^*)_j \; = \; 0. \tag{5.23}$$

Iteration (5.15) becomes

$$y_j^{(k,i)} = y_j^{(k,i-1)} - \lambda_k e^{-y_j{}^{(k,i-1)}} \nabla \hat{L}_i(y^{(k,i-1)})_j, \tag{5.24}$$

for $j = 1, \ldots, n$. This is because, from (5.20), we have

$$\begin{aligned} \nabla \hat{L}_i(y)_j \;&=\; \langle a^i, e^y \rangle \frac{a^i_j e^{y_j}}{\langle a^i, e^y \rangle} + a^i_j e^{y_j} \ln \langle a^i, e^y \rangle - a^i_j e^{y_j} \ln(b_i + 1) \\ &=\; a^i_j e^{y_j} \ln \langle a^i, e^y \rangle - a^i_j e^{y_j} \ln b_i = -a^i_j e^{y_j} \ln \left(\frac{b_i}{\langle a^i, e^y \rangle} \right). \end{aligned}$$

(5.24) means that, for k large enough, $\hat{L}_i$ is decreased at each iteration (see Section 1.2 of [10]).

To simplify notation, we define the diagonal matrix

$$D(y) \;=\; \operatorname{diag}(e^{-y_1}, \ldots, e^{-y_n}), \tag{5.25}$$

and substitute D^k for $D(y^k)$, and in general, we substitute D for $D(y)$, whenever this is clear. Using this notation, (5.24) becomes

$$y^{(k,i)} = y^{(k,i-1)} - \lambda_k D(y^{(k,i-1)}) \nabla \hat{L}_i(y^{(k,i-1)}) \tag{5.26}$$

and since in Subsection 5.2.1, $\{x^{(k,i)}\}$ is bounded, $\{y^{(k,i)}\}$ is bounded from above.

Lemma 5.2.1. *The differences between sub-iterates in (5.24) and major iterates in (5.26) tend to zero, that is, for* $i = 1, \ldots, m$,

$$y^{(k,i)} - y^{(k,i-1)} \underset{k \to \infty}{\longrightarrow} \; 0 \tag{5.27}$$

and

$$y^{k+1} - y^k \underset{k \to \infty}{\longrightarrow} \; 0. \tag{5.28}$$

Also, for $i = 1, \ldots, m$,

$$\frac{x_j^{(k,i)}}{x_j^{(k,i-1)}} \underset{k\to\infty}{\longrightarrow} 1. \tag{5.29}$$

Proof. The sequence $\{x^k\}$ is bounded and so $\{\langle a^i, e^{y^{(k,i-1)}}\rangle\}$ is bounded and

$$\left\{\frac{b_i}{\langle a^i, e^{y^{(k,i-1)}}\rangle}\right\} \tag{5.30}$$

is bounded away from zero for each i. If $\{\langle a^i, e^{y^{(k,i-1)}}\rangle\}$ is bounded away from zero then the factors multiplying the relaxation parameter in (5.15) and (5.17) are bounded and (5.27) and (5.28) hold. If not, then $\langle a^i, e^{y^{(k,i-1)}}\rangle > 0$ for $i = 1, \ldots, m$ and for all $k \geq 0$. This means that (5.30) is not bounded from above and so $\frac{b_i}{\langle a^i, e^{y^{(k,i-1)}}\rangle}$ may tend to $+\infty$ as $k \to \infty$. But, from (5.17), for $j = 1, \ldots, n$, we have

$$y_j^{k+1} = y_j^0 + \sum_{l=0}^{k} \lambda_l \sum_{i=1}^{m} a_j^i \ln\left(\frac{b_i}{\langle a^i, e^{y^{(l,i-1)}}\rangle}\right) \tag{5.31}$$

since $y_j^{k+1} = y_j^{(k,m)}$ and $y_j^k = y_j^{(k,0)}$.

Thus the series on the right hand side of (5.31) may diverge to $+\infty$ due to (5.5) and the fact that (5.30) is not bounded from above. This is a contradiction since $\{y_j^{k+1}\}$ is bounded from above for each j and so $\{\langle a^i, e^{y^{(k,i-1)}}\rangle\}$ must be bounded away from zero. Therefore $\{\langle a^i, e^{y^{(k,i-1)}}\rangle\}$ is bounded and bounded away from zero. Hence (5.29) follows, since $y_j^{(k,i)} - y_j^{(k,i-1)} \longrightarrow 0$ implies $\ln x_j^{(k,i)} - \ln x_j^{(k,i-1)} \longrightarrow 0$ and so $\frac{x_j^{(k,i)}}{x_j^{(k,i-1)}} \longrightarrow 1$, bearing in mind that ln is continuous. $\square$

Another immediate consequence of the rationale in the last lemma is the following corollary. Now, let the set $\mathcal{C}$ be the closure of the convex hull of the iterates $(y^{(k,i)})$ in $\mathcal{R}_-^n$.

Corollary 5.2.2. $\|D(y)\nabla\hat{L}_l(y)\|$ *is bounded in* $\mathcal{C}$.

Proof. Considering only the jth component of $D(y)\nabla\hat{L}_l(y)$, we have

$$\begin{aligned} D(y)_j\nabla\hat{L}_l(y)_j &= e^{-y_j}\nabla\hat{L}_l(y)_j \\ &= e^{-y_j}\left(-a_j^i e^{y_j}\ln\left(\frac{b_i}{\langle a^i, e^{y^{(k,i)}}\rangle}\right)\right) \\ &= -a_j^i \ln\left(\frac{b_i}{\langle a^i, e^{y^{(k,i)}}\rangle}\right) \end{aligned} \tag{5.32}$$

and from the last lemma, $\{\langle a^i, e^{y^{(k,i)}}\rangle\}$ is bounded and bounded away from zero and so $\ln\left(\frac{b_i}{\langle a^i, e^{y^{(k,i)}}\rangle}\right)$ is bounded since ln is continuous and $\mathcal{C}$, being the closure of the convex hull of the bounded iterates $(y^{(k,i)})$ in $\mathcal{R}^n_-$, is bounded. □

The next lemma shows that the sequence is essentially asymptotically decreasing.

Lemma 5.2.3. *If there is an accumulation point $y^* \in \mathcal{R}^n_-$, and a positive number γ such that*

$$\|\nabla \hat{L}(y)\| > \gamma \tag{5.33}$$

for y in an open set Ω in $\mathcal{R}^n_-$ containing y^*, then for k large enough and $y^k \in \Omega$, $\hat{L}(y^{k+1}) \leq \hat{L}(y^k)$. In particular, if (5.33) is valid for every y, then the whole sequence decreases for k large enough.

Proof. Using (5.26) for a complete iteration, we have

$$y^{k+1} = y^k - \lambda_k \sum_{i=1}^{m} (D^{(k,i-1)} \nabla \hat{L}_i(y^{(k,i-1)})) \tag{5.34}$$

and using (5.20), we have

$$D^k \nabla \hat{L}(y^k) = \sum_{i=1}^{m} D^k \nabla \hat{L}_i(y^k). \tag{5.35}$$

Define the algorithm's direction as

$$d^k = \sum_{i=1}^{m} D^{(k,i-1)} \nabla \hat{L}_i(y^{(k,i-1)})$$

so that

$$y^{k+1} = y^k - \lambda_k d^k.$$

Now, using (5.34) and (5.35), for k large enough, we have

$$\begin{aligned} y^{k+1} &= y^k - \lambda_k D^k \nabla \hat{L}(y^k) - \lambda_k \sum_{i=1}^{m} (D^{(k,i-1)} \nabla \hat{L}_i(y^{(k,i-1)}) - D^k \nabla \hat{L}_i(y^k)) \\ &= y^k - \lambda_k D^k \nabla \hat{L}(y^k) + O({\lambda_k}^2), \end{aligned} \tag{5.36}$$

where the last equality is obtained by Lipschitz continuity and boundedness of $D(y)\nabla\hat{L}_i(y)$ on $\mathcal{C}$ by Corollary 5.2.2. That is, for some positive $W \in \mathbb{R}$, we have

$$\left\|\sum_{i=1}^{m} \lambda_k (D^{(k,i-1)}\nabla\hat{L}_i(y^{(k,i-1)}) - D^k\nabla\hat{L}_i(y^k))\right\| \leq \lambda_k W \sum_{i=1}^{m} \left\|y^{(k,i-1)} - y^k\right\|$$

$$= \lambda_k W \sum_{i=1}^{m} \left\|y^{(k,i-1)} - y^{(k,0)}\right\| = \lambda_k W \sum_{i=2}^{m} \left\|\sum_{l=1}^{i-1}(y^{(k,l)} - y^{(k,l-1)})\right\|$$

$$\leq (\lambda_k)^2 W \sum_{i=2}^{m}\sum_{l=1}^{i-1} \left\|D^{(k,l-1)}\nabla\hat{L}_l(y^{(k,l-1)})\right\| \leq (\lambda_k)^2 W \sum_{i=2}^{m}\sum_{l=1}^{i-1} M$$

where

$$M = \max_{y^{(k,l)} \in \mathcal{C}} \left\{\|D^{(k,l-1)}\nabla\hat{L}_l(y^{(k,l-1)})\|\right\}.$$

Therefore

$$\left\|\sum_{i=1}^{m} \lambda_k (D^{(k,i-1)}\nabla\hat{L}_i(y^{(k,i-1)}) - D^k\nabla\hat{L}_i(y^k))\right\| \leq (\lambda_k)^2 m^2 W M$$

and so for k large enough,

$$-\lambda_k \sum_{i=1}^{m}(D^{(k,i-1)}\nabla\hat{L}_i(y^{(k,i-1)}) - D^k\nabla\hat{L}_i(y^k)) = O(\lambda_k{}^2).$$

Now consider the objective function $\hat{L}(y^k)$. Since $\nabla\hat{L}_i(y)$ is Lipschitz continuous on $\mathcal{C}$, for some $L_0 > 0$, we have from [[77], precisely (15) on p.6],

$$\left\|\hat{L}(y^{k+1}) - \hat{L}(y^k) - (y^{k+1} - y^k)^T\nabla\hat{L}(y^k)\right\| \leq \frac{L_0}{2}\|y^{k+1} - y^k\|^2.$$

Therefore, from (5.28), for k large enough, we have

$$\hat{L}(y^{k+1}) = \hat{L}(y^k) + (y^{k+1} - y^k)^T\nabla\hat{L}(y^k) + O(\|y^{k+1} - y^k\|^2). \tag{5.37}$$

Now, using (5.36) for k large enough and the fact that $D^k\nabla\hat{L}(y^k)$ is bounded , we have

$$\begin{aligned}(y^{k+1} - y^k)^T\nabla\hat{L}(y^k) &= (-\lambda_k D^k\nabla\hat{L}(y^k) + O(\lambda_k{}^2))^T\nabla\hat{L}(y^k) \\ &= -\lambda_k(\nabla\hat{L}(y^k))^T D^k\nabla\hat{L}(y^k) + O(\lambda_k{}^2)\nabla\hat{L}(y^k) \\ &= -\lambda_k(\nabla\hat{L}(y^k))^T D^k\nabla\hat{L}(y^k) + O(\lambda_k{}^2)\end{aligned}$$

and

$$\begin{aligned}\|y^{k+1} - y^k\|^2 &= (-\lambda_k D^k \nabla \hat{L}(y^k) + O({\lambda_k}^2))^T(-\lambda_k D^k \nabla \hat{L}(y^k) + O({\lambda_k}^2)) \\ &= {\lambda_k}^2 (D^k \nabla \hat{L}(y^k))^T D^k \nabla \hat{L}(y^k) + O({\lambda_k}^2)\lambda_k D^k \nabla \hat{L}(y^k) \\ &\quad + O({\lambda_k}^2) O({\lambda_k}^2) \\ &= O({\lambda_k}^2) + O({\lambda_k}^2) + O({\lambda_k}^2) = O({\lambda_k}^2),\end{aligned}$$

due to (5.4). Therefore from (5.37), for k large enough, we have

$$\hat{L}(y^{k+1}) = \hat{L}(y^k) - \lambda_k (\nabla \hat{L}(y^k))^T D^k \nabla \hat{L}(y^k) + O({\lambda_k}^2) \tag{5.38}$$

and the result holds since the factor multiplying λ_k is bounded away from zero. This is because D^k is bounded away from zero, since in the proof of Lemma 5.2.1, x^k or e^{y^k} is bounded away from zero. Thus, using (5.33), $(\nabla \hat{L}(y^k))^T D^k \nabla \hat{L}(y^k)$ is bounded away from zero. □

5.2.3 Limit points

In the next proposition, we further assume that in addition to conditions (5.4) and (5.5), the relaxation parameter λ_k is chosen such that $\sum_{k=1}^{\infty} (\lambda_k)^2 < \infty$. In this case, $\sum_{k=1}^{\infty} O(\lambda_k)^2 < \infty$.

Proposition 5.2.4. *If $\sum_{k=1}^{\infty} (\lambda_k)^2 < \infty$ then there is a limit point x^* of the sequence (5.6) such that*

$$x_j^* \nabla L(x^*)_j = 0. \tag{5.39}$$

Equivalently

$$\nabla \hat{L}(y^*) = 0, \tag{5.40}$$

where $y^* \in \mathcal{R}_-^n$ and $y_j^* = \ln x_j^*$ for $j = 1, \ldots, n$.

Proof. If there is no such a limit point then using (5.22),

$$D^{-1}(y^k) D(y^k) \nabla L(x^k) = D(y^k) \nabla \hat{L}(y^k)$$

is bounded away from zero and there exists a real positive number η such that

$$(\nabla \hat{L}(y^k))^T D^k \nabla \hat{L}(y^k) > \eta > 0.$$

Thus, by Lemma 5.2.3, $\hat{L}(y^{k+1}) \leq \hat{L}(y^k)$ for every $k \geq k_0$, and, for every integer $l > k_0$, we have

$$\hat{L}(y^{l+1}) - \hat{L}(y^{k_0}) = \sum_{k=k_0}^{l} (\hat{L}(y^{k+1}) - \hat{L}(y^k)) \leq -\eta \sum_{k=k_0}^{l} \lambda_k + \sum_{k=k_0}^{l} O({\lambda_k}^2). \quad (5.41)$$

But, since by hypothesis, $\lim_{l\to\infty} \sum_{k=k_0}^{l} O({\lambda_k}^2)$ is finite, the right hand side of (5.41) tends to $-\infty$ as l tends to infinity due to (5.5) whiles the left hand side is bounded which is a contradiction. □

Theorem 5.2.5. *Every limit point of the sequence generated by (5.6) satisfies (5.39).*

Proof. From Proposition 5.2.4, we know that such a limit point, say x^* or y^*, exists. Suppose that there is another limit point $x^{**} = \lim_{j\to\infty} x^{k_j}$ ($y^{**} = \lim_{j\to\infty} y^{k_j} = \lim_{j\to\infty} \ln x^{k_j}$, such that $x^{**}\nabla L(x^{**}) \neq 0$ or $\nabla L(y^{**}) \neq 0$). Then, it is clear from (5.23) that, $\lim_{j\to\infty} \nabla\hat{L}(y^{k_j}) \neq 0$, that is, $\|\nabla\hat{L}(y^{k_j})\| > \gamma$ for some positive γ.

Now let LP be the set of limit points of the sequence $\{y^k\}$ such that $\nabla L(y) = 0$ or $x\nabla L(x) = 0$. Then LP is closed and bounded since the sequence is bounded. For a given positive ϵ, define the set $LP_\epsilon := \{y \mid \text{distance}(y, LP) \leq \epsilon\}$, and assume that the limit point $y^{**} \notin LP_\epsilon$ and that distance$(y^{**}, LP_\epsilon) = \delta > 0$. Then $\|\nabla\hat{L}(y^{k_j})\| > \gamma$ for $y^{k_j} \in \Omega$, where Ω is some open neighbourhood of y^{**} in $\mathcal{R}^n_-$ and a complement of LP_ϵ in $\mathcal{R}^n_-$.

Now considering the whole sequence $\{y^k\}$, by Lemma 5.2.3, $\{\hat{L}(y^k)\}$ decreases for k large enough and $y^k \in \Omega$, i.e., there exists a constant k_0 such that $\hat{L}(y^{k+1}) \leq \hat{L}(y^k)$ for $k \geq k_0$ if $y^k \notin LP_\epsilon$ or $y^k \in \Omega$. But, since there is a subsequence converging to $y^* \in LP$, there exists $k_1 > k_0$ such that $y^{k_1} \in LP_\epsilon$ and $\hat{L}(y^*) < \hat{L}(y^k)$. This is because if $\hat{L}(y^*) \leq \hat{L}(y^k)$ for k large enough and $y^k \in LP_\epsilon$ then

(i) there could not exist a subsequence converging to y^{**} this is because the distance between iterates tends to zero, i.e., $y^{k+1} - y^k \longrightarrow 0$,

(ii) the sequence $\{\hat{L}(y^k)\}$ decreases outside LP_ϵ and the distance between y^{**} and LP_ϵ is positive.

Thus, since $y^{k_1} \in LP_\epsilon$, if by continuity $LP_\epsilon \longrightarrow LP$ as $\epsilon \longrightarrow 0$ then y^{k_1} tends to say $\bar{y}$ as ϵ tends to zero (or a convergent subsequence for ϵ tending to zero). Therefore $\bar{y} \in LP$, $\nabla \hat{L}(\bar{y}) = 0$ and $\hat{L}(\bar{y}) \geq \hat{L}(y^*)$ which is a contradiction.

□

5.2.4 Convergence of the whole sequence

The next proposition is trivial, reflecting the fact that the solution set is the intersection of a linear system with the nonnegative orthant. The proof can be found in [19].

Proposition 5.2.6. *If x^* solves (5.7) then every other solution x (nonnegative) solves the linear system $Ax = Ax^*$.*

Theorem 5.2.7. *The whole sequence generated by (5.6) converges to a maximum entropy (relative to the starting point) solution of (5.7).*

Proof. From (5.16), it is clear that

$$\nabla K(x^k, \mathbf{1}) = \nabla K(x^{k-1}, \mathbf{1}) + A^T z^{k-1} \tag{5.42}$$

for some vector $z^{k-1} \in \mathbb{R}^m$ with $z_i^{k-1} = \lambda_k \ln\left(\frac{b_i}{\langle a^i, x^{(k,i-1)}\rangle}\right)$ and $\mathbf{1}$ stands for the vector of ones. Therefore iterating, we have

$$\nabla K(x^k, \mathbf{1}) - \nabla K(x^0, \mathbf{1}) = \nabla K(x^k, x^0) = A^T \sum_{s=1}^{k} z^{k-s} \tag{5.43}$$

and taking limits, we have

$$\nabla K(x^*, x^0) = 0 \in \mathrm{Im}(A^T) \tag{5.44}$$

since, in this case, for $j = 1, \ldots, n$,

$$\lim_{k\to\infty} \sum_{s=1}^{k} \sum_{i=1}^{m} a_j^i z_i^{k-s} = \lim_{k\to\infty} \sum_{s=1}^{k} \lambda_s \sum_{i=1}^{m} a_j^i \ln\left(\frac{b_i}{\langle a^i, x^{(s,i-1)}\rangle}\right) = 0$$

in view of (5.5) and the fact that x^* is a maximum entropy solution of (5.7). We must therefore have

$$\lim_{s\to\infty}\sum_{i=1}^{m} a_j^i \ln\left(\frac{b_i}{\langle a^i, x^{(s,i-1)}\rangle}\right) = \sum_{i=1}^{m} a_j^i \ln\left(\frac{b_i}{\langle a^i, x^*\rangle}\right) = 0.$$

But, considering Proposition 5.2.6, (5.44) completes the Kuhn-Tucker conditions for the problem

$$\min_{x\geq 0}\ K(x, x^0) \text{ subject to } x = \arg\min_{x\geq 0} L(\hat{x}). \tag{5.45}$$

The problem (5.45) has a unique solution (entropy is a strictly convex function with linear constraints); so, the whole sequence converges. □

5.3 On SMART

In this section, we deduce the Simultaneous Multiplicative Algebraic Reconstruction Technique, better known as SMART (see [18]), as a particular case of the majorizing functions technique, introduced by De Pierro in [15, 39], applied to the function defined by (5.7). Writing $\langle a^i, x\rangle = \sum_{j=1}^{n}\left(\frac{a_j^i x_j^k}{\langle a^i, x^k\rangle}(\frac{x_j}{x_j^k})\langle a^i, x^k\rangle\right)$ and noting that $\sum_{j=1}^{n}\frac{a_j^i x_j^k}{\langle a^i, x^k\rangle} = 1$, and using convexity of the $x\ln x$ functional, we have the following inequality,

$$L(x) \leq \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{a_j^i x_j^k}{\langle a^i, x^k\rangle}\cdot\frac{x_j}{x_j^k}\langle a^i, x^k\rangle \ln\frac{x_j}{x_j^k}\langle a^i, x^k\rangle - (1+\ln b_i)\langle a^i, x\rangle\right), \tag{5.46}$$

i.e.,

$$L(x) \leq \varphi(x, x^k) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(a_j^i x_j \ln\frac{x_j}{x_j^k}\langle a^i, x^k\rangle - (1+\ln b_i)\langle a^i, x\rangle\right). \tag{5.47}$$

The function $\varphi(x, x^k)$ is strictly convex with separated variables and has the following properties:

(i) $\varphi(x^k, x^k) = L(x^k)$,

(ii) $\nabla_x\varphi(x^k, x^k) = \nabla L(x^k)$.

Essentially this means that the new function majorizes the one that we want to minimize and coincides with its gradient at the current iterate x^k.

The majorizing algorithm is then defined iteratively as:

$$x^{k+1} = \arg\min_{x\geq 0}\ \varphi(x, x^k). \tag{5.48}$$

In order to compute the next iterate, we differentiate the variable separable function $\varphi(x, x^k)$ and have

$$\frac{\partial \varphi(x, x^k)}{\partial x_j} = \sum_{i=1}^{m} a_j^i \left(-\ln b_i + \ln x_j \frac{\langle a^i, x^k \rangle}{x_j^k} \right) = \ln \prod_{i=1}^{m} \left(\frac{x_j}{x_j^k} \frac{\langle a^i, x^k \rangle}{b_i} \right)^{a_j^i} = 0. \tag{5.49}$$

From this equation, we have

$$(x^{k+1})^{\sum_i a_j^i} = (x^k)^{\sum_i a_j^i} \prod_{i=1}^{m} \left(\frac{b_i}{\langle a^i, x^k \rangle} \right)^{a_j^i}. \tag{5.50}$$

When the row sums, $\sum_i a_j^i$, are one, we recover SMART. Following this approach, a convergence proof of SMART can be obtained, but this has been done in a more general setting by C. Byrne in [18].

Chapter 6

New results, conclusion and future work

6.1 New results and conclusion

We presented new results on Bregman iterative methods. In Chapter 2, a new approach to Bregman projection methods for the convex feasibility problem that generalizes the concept of relaxation introduced in [41, 29] was presented. This generalization allowed us to define approximate Bregman's projection method for general not necessarily linear convex feasibility problems. As a consequence, we derived an application to convex but nonlinear sets of constraints and to linear equality constraints as well.

In Chapter 3, we discussed block type methods for linear inequality constrained problems and the corresponding convergence proofs emphasizing on the dual approach. To the best of our knowledge, the simultaneous version of the Bregman's algorithmic scheme for solving linearly constrained optimization problem formulated in Chapter 3 does not exist anywhere in the literature. Therefore our simultaneous algorithmic scheme is a new method for minimizing any Bregman function over linear constraints. A convergence proof of this algorithm was also given.

Again, for the first time, a general closed-form formula for the iterative step in

Bregman's algorithm for the optimization of any Bregman function with separated variables was presented in Chapter 4. The secant approximation approach used in [23] to establish the relationship between MART and Entropy maximization over linear equalities was the motivation.

In Chapter 5, we analyzed the behaviour of Bregman type algorithm when the problem is inconsistent. We have proven that adequately underrelaxed MART for equalities converges to a minimum solution of Kullback Leibler distance.

6.2 Future work

In future, we will illustrate the conjecture outlined at the end of Chapter 2 by some numerical examples and prove it. This will enable us construct an algorithm using a sequence of Bregman projections in the general case, i.e., where the constraints are not only linear equalities, to minimize the Bregman function involved. We will then examine the specific case where the Bregman function is quadratic for norm minimization.

In Chapter 3, we produced a simultaneous Bregman's algorithm that minimizes Bregman function over polyhedron. As part of our future work, we will consider the case where the polyhedron is empty and find an approximate solution for the minimization problem. This will be a generalization of the result in [40], i.e., the simultaneous Hildreth method where the function $\frac{1}{2}\|\cdot\|^2$ was minimized over the empty polyhedron.

We will also consider a generalization of the work in [73], as a consequence of a result in [2]. We will do this by proving the following statements:

(i) Any sequence of orthogonal projections onto a finite family of polyhedral convex sets is bounded.

(ii) Any sequence of Bregman projections (including the orthogonal projections) onto a finite family of affine sets is bounded.

(iii) Any sequence of Bregman projections onto a finite family of polyhedral convex sets is bounded.

The Algorithmic schemes 4.1.1 and 4.1.2 proposed in Chapter 4 are without convergence results. To establish the practicability of these algorithmic schemes, we shall first prove the convergence results of the schemes and demonstrate numerically the capabilities of these schemes for a particular Bregman function, i.e., the negative Shannon entropy as defined by the iterative steps (4.12) for solving large scale problems in image reconstruction from projections. We will go ahead and compare numerically the performance of the algorithm MART and the algorithm for entropy maximization over linear equalities as defined in (4.12). It will also enable us to compare the efficiencies of our proposed algorithms for maximization of Shannon entropy with that of the existing similar algorithms in the literature [22].

Finally, we will work on a MATLAB package with implementation of MART and SMART algorithms for discretization of inverse problems. In this regard, we shall provide a few simplified 2D and 3D tomography test problems from x-ray CT and then undertake a series of experiments to determine the importance of these algorithms with underrelaxation to image quality.

Bibliography

[1] Agmon S. 1954, The Relaxation Method for Linear Inequalities, *Canad. J. Math*, Vol. 6, pp. 382-392.

[2] Aharoni R., Duchet P. and Wajnryb B. 1984, Successive Projections on Hyperplanes, *Journal of Mathematical Analysis and Applications*, Vol. 103, pp. 134-138.

[3] Aharoni R., Breman A. and Censor Y 1983, An Interior Points Algorithm for the Convex Feasibility Problem, *Advances in Applied Mathematics*, Vol. 4, pp. 479-489.

[4] Aharoni R., Censor Y., De Pierro A. R. and Zaknoon M. 2004, On the convergence of sequential projection algorithms for the inconsistent convex feasibility problem, submitted.

[5] Aharoni R. and Censor Y. 1989, Block-iterative projection methods for parallel computation of solutions to convex feasibility problems, *Linear Algebra and its Applications*, Vol. 120, pp. 165-175.

[6] Avriel M. 1976, Nonlinear Programming: Analysis and Methods, *Prentice Hall, Englewood Cliffs, NJ.*

[7] Bauschke H. H. and Combettes P. L. 2011, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, *Canadian Mathematical Society*, Address: Halifax, Nova Scotia, Canada.

[8] Beyer W. A. and Ulam S. M. 1968, Note on the visual hull of a set, *J. Comb.*

Theory, Vol. 4, 240-245.

[9] Bertsekas D. P. and Tsitsiklis J. N. 1996, Neuro-Dynamic Programming, *Athena Scientific, Bel-mont, MA.*, address: Athena Scientific, Belmont Mass. U.S.A

[10] Bertsekas D. P. 2003, Nonlinear Programming *Athena Scientific, Belmont, Massachusetts*, address: Athena Scientific, Belmont Mass. U.S.A

[11] Bertsekas D. P. 2003, Convex Analysis and Optimization, *Athena Scientific, Belmont, Massachusetts.*, address: Athena Scientific, Belmont Mass. U.S.A

[12] Boyd S. and Vandenberghe L. 2009, Convex Optimization, *Cambridge University Press*, Address: The Edinburgh Building, Cambridge, UK.

[13] Bregman L. M., Censor Y. and Reich S. 1999, Dykstra's Algorithim as the Nonlinear Extension of Bregman's Optimization Method, *J. Convex Analysis*, Vol. 6, pp. 319-333.

[14] 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. 'Com-put.' Math. and Math. Phys,*, Vol. 7, pp. 200-217.

[15] Browne J. A. and De Pierro A. R. 1996, A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography, *IEEE Transactions on Medical Imaging*, Vol. 15, pp. 687-699.

[16] Burg J. P. 1967, Maximum Entropy Spectral Analysis, *Proceedings 37th Annual Meeting of the Society of Exploration Geophysicists, Oklahoma City, OK.*

[17] Burg J. P. 1967, Maximum Entropy Spectral Analysis, *Proceedings 37th Annual Meeting of the Society of Exploration Geophysicists*, Vol. 3

[18] Byrne C. 1996, Block iterative methods for image reconstruction from projections, *IEEE Trans. Image Process.* IP-5, 792-794.

[19] Byrne C. 1994, Iterative Reconstruction Algorithms Based on Cross-Entropy Minimization, in Image Models (and their Speech Model Cousins), S. E. Levinson and L. Shepp, Editors, IMA Volumes in Mathematics and its Applications, Vol. 80, pp. 1-11 New York: Springer-Verlag.

[20] Censor Y. 1979, Row-action methods for huge and sparse systems and their applications, *Tech. Rep. MIPG 28, Medical Image Processing Group, Dept. of Computer Science, State Univ. of New York.*

[21] Censor Y. 1981, "Row-action methods for huge and sparse systems and their applications", SIAM Rev., Vol. 23, pp. 444-466.

[22] Censor Y., De Pierro A. R. and Iusem A. N. 1991, Optimization of Burg's Entropy over Linear Constraints, *Applied Numerical Mathematics*, Vol. 7, pp. 151-165.

[23] Censor Y., De Pierro A. R., Elfving T., Herman G.T, and Iusem A. N. 1990, On Iterative Methods for Linearly Constrained Entropy Maximization, *Numerical Analysis and Mathematical Modelling, Banach Center Publications*, Vol. 24, pp. 145-163.

[24] Censor Y., De Pierro A. R. and Zaknoon M. 2004, Steered Sequential Projections for the Inconsistent Convex Feasibility Problem, *Transportation Res.*, Vol. 59, pp. 385-405.

[25] Censor Y., Eggermont P. P. B. and Gordon D. 1983, Strong underrelaxation in Kaczmarz's method for inconsistent systems, *Numer. Math.* Vol. 41, pp. 83-92.

[26] Censor Y., Elfving T. and Herman G.T. 1977, Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems, *Special-purpose algorithms for linearly constrained entropy maximization, in: C.R. Smith and G.J. Erickson, eds.* , pp.241-254.

[27] Censor Y., Elfving T. and Herman G.T. 1983, Methods for Entropy Maxi-

mization with Applications in Image Processing, in: P. Johansen and P.W. Becker, eds., *Proceedings Third Scandinavian Conference on Image Analysis* , pp. 296-300.

[28] Censor Y. and Herman G.T. 1979, Row-generation methods for feasibility and optimization problems involving sparse matrices and their applications, *Sparse Matrix Proceedings 1978, I.S. Duff and G. W. Stewart, eds., Society for Industrial and Applied Mathematics, Philadelphia*, Vol. 4, pp. 197-219.

[29] Censor Y. and Herman G. T. 2002, Block-Iterative Algorithms with Under-relaxed Bregman Projections, *SIAM Journal on Optimization*, Vol. 13, pp. 283-297.

[30] Censor Y. and Lent A. 1981, An Iterative Row-action Method for interval convex programming, *Journal of optimization Theory and Applications*, Vol. 34, pp. 321-353.

[31] Censor Y. and Segmen J. 1987, On block-iterative Entropy Maximization *Journal of Information and Optimization Sciences*, Vol. 8, pp. 275-291.

[32] Censor Y. and Zenios S. A. 1997, Parallel Optimization; Theory, Algorithms, and Applications, *Oxford University Press, Inc.*, address: New York, USA.

[33] Chaudhuri B. B. and Rosenfeld A. 1998, On the computation of the digital convex hull and circular hull of a digital region, *Pattern Recognition*, Vol. 31, 2007-16.

[34] Cimmino G. 1938, Calcoho approssimato per le soluzioni dei sistemi di equazioni lineari, *La Ricerca Scientifica, Roma*, Vol. XVI, Ser. II. Anno IX, pp. 326-333.

[35] Csiszàr I. and Tusnády G. 1984, Information geometry and alternating minimization procedures', in Statistics and Decisions, Supplement Issue No. 1, R. Oldenbourg Verlag, pp. 205-237.

[36] Daniel J. W. 1971, The Approximate Minimization of Functionals, *Prentice-*

Hall, Address:Englewood Cliffs, NJ.

[37] Darroch J. N. and Ratcliff D. 1972, Generalized iterative scaling for log-linear models, *Annals of Mathematical Statistics*, Vol. 43, 1470-80.

[38] Dempster A. P., Laird N. M. and Rubin D. D. 1977, Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Stat. Soc.*, Series B, Vol. 39, pp. 1-38.

[39] De Pierro A. R. 1995, A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography, *IEEE Trans. Med. Imaging*, Vol. 14, 1, pp. 132-137.

[40] De Pierro A. R. and Iusem A. N. 1985, "A simultaneous projection method for linear inequalities", *Linear Algebra and its Applications*, Vol. 64, 243-253.

[41] De Pierro A. R. and Iusem A. N. 1986, A Relaxed Version of Bregman's Method for Convex Programming, *Optimization Theory and its Applications*, Vol. 5, pp. 421-440.

[42] De Pierro A. R. and Iusem A. N. 1988, A Finitely Convergent "'Row-Action"' Method for the Convex Feasibility Problem, *Applied Mathematics and Optimization*, Vol. 17, pp. 225-235.

[43] De Pierro A. R. and Iusem A. N. 1989, On the Asymptotic Behavior of Some Alternate Smoothing Series Expansion Iterative Methods, *Applied Numerical Mathematics*, Vol. 29, pp. 421-440.

[44] De Pierro A. R. and Iusem A. N. 1990, A Finitely Convergent "'Row-Action"'Method for the Convex Feasibility Problem, *Mathematical Programming*, Vol. 47, pp. 37-51.

[45] De Pierro A. R. and Iusem A. N. 1991, On The Convergence of Han's Method for Convex Programming with Quadratic Objective, *Mathematical Programming*, Vol. 52, pp. 265-284.

[46] D'Addario L. R. 1977, Maximum-Entropy Imaging, *Theory and philosophy, in: R. Shaw, ed.,Image Analysis and Evaluation (Society of Photographic Scientists and Engineers(SPSE))*, pp. 221-225.

[47] Elfving T. 1980, One Some Methods for Entropy Maximization and Matrix Scaling, *Linear Algebra and its Applications*, Vol. 34, pp. 329-343.

[48] Erlander S. 1981, Entropy in Linear Programmes, *Mathematical Programming*, Vol. 21, pp. 137-151.

[49] Eggermont P.P.B., Herman G.T. and Lent A. 1981, Iterative Algorithms for Large Partitioned Linear Systems, with Applications to Image Reconstruction, *Linear Algebra Appl.*, Vol. 40, pp. 37-67.

[50] Frieden B.R. 1980, Statistical Models for the Image Restoration Problem, *Comput. Graph Image Process*, Vol. 12, pp. 40-59.

[51] Gordon R., Bender R. and Herman G. T. 1970, Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and x-ray photography, *J. Theor. Biol.*, Vol. 29, 471-81.

[52] Gordon R. and Herman G. T. 1974, Three dimensional reconstruction from projections: a review of algorithms, *Int. Rev. Cytol.*, Vol. 38, 111-51.

[53] Gordon R, Bender R. and Herman G. T. 1970, Algebraic reconstruction techniques (ART) for three dimensional electron microscopy and x-ray photography, *Journal of Theoretical Biology*, Vol. 29, pp. 471-481.

[54] Gubin L., Polyak B. and Raik E. 1967, The Method of Projections for Finding the Common Point of Convex Sets, *USSR Computational Mathematics and Mathematical Physics*, Vol. 7, pp. 1-24.

[55] Herman G. T. 1980, Image Reconstruction from Projections: The Fundamentals of Computerized Tomography, *Academic Press, New York* .

[56] Herman G.T. and Lent A. 1978, A Family of Iterative Quadratic Optimiza-

tion Algorithms for Pairs of Inequalities, with Applications in Diagnostic Radiology, *Mathematical Programming Study*, Vol. 9, pp. 15-29.

[57] Herman G. T. and Lent A. 1976, Iterative Reconstruction Algorithms, *Computers in Biology and Medicine*, Vol. 6, pp. 273-294.

[58] Hildreth C. 1957, Quadratic Programming Procedure, *Michigan Agricultural Experiment Station*, Issue No. 361, Vol. 4, pp. 79-85.

[59] Hoffman A. J. 1952, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, *J. Res. Nat. Bur. Standards*, A Vol. 49, pp. 263-265.

[60] Hounsfield G. N. 1973, Computerized transverse axial scanning tomography, Part I: Description of the system, *British Journal of Radiology*, Vol. 46, pp. 1016-1022.

[61] Iusem A. N. and Svaiter B. F. 1994, A Row-Action Method for Convex Programming *Mathematical Programming*, Vol. 64, pp. 149-171.

[62] Iusem A. N. and De Pierro A. R. 1987, A Simultaneous Iterative Method for Computing Projections on Polyhedra, *Siam J. Control and Optimization*, Vol. 25, pp. 231-242.

[63] Iusem A. N. and De Pierro A. R. 1990, On The Convergence Properties of Hildreth's Quadratic Programming Algorithm, *Discrete Mathematics*, Vol. 47, pp. 37-51.

[64] Iusem A. N. and De Pierro A. R. 1991, Optimization of Burg's Entropy over Linear Constraints, *Applied Numerical Mathematics*, Vol. 46, pp. 265-284.

[65] Jaynes E. T. 1982, On the rational of maximum entropy methods, *Proceedings of the IEEE*, Vol. 70, pp. 939-952.

[66] Johnson R. and Shore J. E. 1984, Which is the better Entropy Expression for Speech Processing:-SlogS or logS, *IEEE Trans. Acoust. Speech Signal Process*, Vol. 32, pp. 129-136.

[67] Kaczmarz S. 1937, Angenaherte Auflosung von Systemen linearer Gleichungen, *Bull. Acad. Polon. Sci. Lett.*, Vol. 35, pp. 355-357.

[68] Kapur J. N. 1983, A Comparative Assessment of Various Measures of Entropy, *J. Information Optimization Sciences*, Vol. 4, pp. 207-232.

[69] Kullback S. and Leibler R., 1951, "On information and sufficiency", *Ann. Math. Stat.* Vol. 22, 1, pp. 79-86.

[70] Lamond B. and Stewart N. P. 1981, Statistical Models for the Image Restoration Problem, *Transportation Res.*, Vol. 15B, pp. 239-248.

[71] Lent A. 1977, A convergent algorithm for maximum entropy image restoration with a medical x-ray application", In: Shaw, *Image Analysis and Evaluation*, pp. 249-257, Washington, D.C., Society of Photographic Scientists and Engineers.

[72] Lent A. and Censor Y. 1980, Extensions of Hildreth's Row-Action Method For Quadratic Programming, *Siam J. Control and Optimization*, Vol. 18, pp. 444-454.

[73] Meshulam R. 1996, On Products of Projections, *Discrete Mathematics*, Vol. 154, pp. 307-310.

[74] Nashed M. Z. 1981, Continuous and semicontinuous analogues of iterative methods of Cimmino and Kaczmarz with applications to inverse Radon transform, Mathematical Aspects of Computerized Tomography, *Lectures Notes in Medical Informatics, Springer Verlag Berlin*, Editors:Herman G. T. and Natterer F. Vol. 8, 160-178.

[75] Nowak R. D. 2003, Distributed EM algorithms for density estimation and clustering in sensor networks, *IEEE Trans. Signal Process.*, Vol. 51, pp. 2245-2253.

[76] Ostrowski A. M. 1973, *Solution of Equations in Euclidean and Banach Spaces*, New York, Academic Press.

[77] Polyak B. T. 1987, Introduction to Optimization, *New York: Optimization Software.*

[78] Rockafellar R. T. 1970, Convex Analysis, *Princeton University Press.*, address: Princeton, New Jersey, USA

[79] Shannon C. E. 1948, A Mathematical Theory of Communication, *Bell System Technical Journal*, Vol. 27, pp. 379-423.

[80] Solodov M. V. 2003, Incremental gradient algorithms with stepsizes bounded away from zero, *"'Com-put."' Optim. Appl.*, Vol. 11, pp. 23-35.

[81] Trummer M. R. 1984, Short Communications / Kurze Mitteilungen, A Note on the ART of Relaxation, *Computing by Springer-Verlag*, Vol. 33, pp. 349-352.

[82] -,-. 1986, A relaxed version of Bregman's method for convex programming, *J. Optim. Theory and Application*, Vol. 51, 421–440.