OPTIMIZATION OF BURG'S ENTROPY OVER LINEAR CONSTRAINTS

Yair Censor
Alvaro R. De Pierro
and
Alfredo N. Iusem

RELATÓRIO TÉCNICO Nº 27/89

Abstract. Two special–purpose iterative algorithms for maximization of Burg's entropy function subject to linear inequalities are presented. Both are "row–action" methods which use in each iteration the information contained in only one constraint. One is an under–relaxed Bregman's algorithm which requires, at each iterative step, the solution of a system of equations. In contrast with this, the second algorithm employs a closed–form formula for the iterative step. Complete analyses of the convergence for both algorithms are given.

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Ciência da Computação
IMECC – UNICAMP
Caixa Postal 6065
13.081 – Campinas – SP
BRASIL

Agosto – 1989

1. INTRODUCTION

The problem of maximizing an entropy function over linear constraints (equality, inequality or interval constraints), arises in many fields of applications. These include: transportation planning (see, e.g., [29]), statistics (see, e.g., [14]), linear numerical analysis (see, e.g., [16]), chemistry (see, e.g., [17]), geometric programming (see, e.g., [32]), image reconstruction from projections (see, e.g., [5,21]), image restoration (see, e.g. [19]), pattern recognition (see, e.g., [1]) and spectral analysis (see, e.g., [2]). Further information on the use of entropy maximization in these and in other fields may be found in [20,25,27].

A large number of measures of entropy have been proposed in different fields of applications, see [28]. We are concerned here with two of them. The first one, Burg's entropy, $B(x)$, also known as "log x-entropy", is defined on $\mathbb{R}^n_>$ the positive orthant of the n-dimensional Euclidean space $\mathbb{R}^n$ by:

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

We also refer to the "x log x-entropy", ent x, also known as Shannon's entropy, defined on the nonnegative orthant $\mathbb{R}^n_{\geq}$ by:

$$\text{ent } x = - \sum_{j=1}^{n} x_j \log x_j \, , \qquad (2)$$

where, by convention $0 \log 0 = 0$.

Burg's entropy was first proposed in [2] (see also [18, sections 5.17, 5.18] and [25]) 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 [13,18,31,20 (Section 10.4.14)] and recently also in [26].

We are interested in "row-action" methods for maximization of entropy functions subject to linear constraints. These methods do not modify the matrix of the contraints and use only one constraint at a time as iterations progress. As such, they are advantagous for handling large and sparse systems. See [4] for a discussion of "row-action" methods and applications for which they are particularly suitable.

For Shannon's entropy function, two well-known "row-action" methods are available. One is Bregman's method [3] which applies in fact for a larger set of objective functions, called "Bregman functions" in [11]. The other one is MART ("Multiplicative Algebraic Reconstruction Technique"), which has been used in the fields of Transportation Research [29], Image Reconstruction from Projections [4,5,30] and others. MART for inequalities has been studied in [8,9].

Introduction of underrelaxation in Bregman's algorithm [15] made it possible to consider MART as an instance of Bregman's algorithm with a special underrelaxation strategy, thus clarifying the relation between the two algorithms and providing a more comprehensive convergence analysis for MART [6].

The advantage of MART over Bregman's method is that the latter requires in general the solution of a non-linear equation in one unknown at each iteration, while MART substitutes a closed formula for that equation (i.e., the relaxation parameters are chosen in such a way that the non-linear equation has an analytical solution).

In this paper we extend this approach to Burg's entropy function. In the case of equality constraints Bregman's algorithm for this function was proposed in [10] without any convergence analysis and implemented in [24] for a problem of two-dimensional spectral estimation. An ad-hoc covergence proof was given in [12]. We present here a version of Bregman's method for the maximization of Burg's entropy function subject to linear inequality constraints. Besides the nature of the constraints (equality vs. inequality) this algorithm differs from the method in [10,12] in that it includes built-in relaxation, as in [15], and allows more freedom in the control strategy (i.e., the order in which the constraints are utilized).

We present also a "MART-type" algorithm for maximization of Burg's entropy subject to linear inequality constraints (which we call the "hybrid algorithm"). As in the case of MART, the improvement over Bregman's method lies in the fact that no inside loop of numerical solution of non-linear equations is required.

Convergence of Bregman's algorithm for Burg's entropy is not an immediate consequence of the convergence theorems of [15] because Burg's entropy is not, strictly speaking, a Bregman function. Bregman functions must be, according to [2,11,15], continuous at the boundary of their domains. Burg's entropy is defined in the interior of the nonnegative orthant of $\mathbb{R}^n$ and is not continuous on its boundary. A technical adjustment in the convergence proof is therefore performed to circumvent this obstacle. This technique is not exclusive for Burg's entropy; it has been applied in [7] in order to extend Bregman's convergence analysis to arbitrary Bregman functions which have singularities on the boundary of their domains.

Convergence of the "hybrid algorithm" is proved by showing that it is obtained as a result of a particular choice of the relaxation parameters in Bregman's algorithm.

2. BREGMAN'S ALGORITHM

Bregman's algorithm is designed to solve linearly constrained optimization problems of the form

$$\min f(x) \quad \text{s.t.} \quad Ax \le b \,, \tag{3}$$

with $f: \Lambda \subseteq \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.

Before statement of the algorithm and of the conditions on f necessary for its convergence, some definitions are required. Associate with f a non-empty, convex, open set $S \subseteq \Lambda$ (called the *zone* of f). $\bar{S}$ will denote the closure of S.

Define the function $D: \bar{S} \times S \to \mathbb{R}$ by

$$D(x,y) = f(x) - f(y) - \langle \nabla f(y), x-y \rangle , \tag{4}$$

where ∇ denotes the gradient and $\langle \cdot , \cdot \rangle$ the Euclidean inner product in $\mathbb{R}^n$.

Finally consider the level sets of D for $\gamma \in \mathbb{R}$:

$$L_1(y,\gamma) = \left\{ x \in \bar{S} \mid D(x,y) \le \gamma \right\} ,$$
$$L_2(x,\gamma) = \left\{ y \in S \mid D(x,y) \le \gamma \right\} . \tag{5}$$

The conditions on f required for convergence of the algorithm are:

(*i*) f is continuously differentiable on S,

(*ii*) f is strictly convex and continuous on $\bar{S}$,

(*iii*) For every $\gamma \in \mathbb{R}$ the sets $L_1(y,\gamma)$, $L_2(x,\gamma)$ are bounded, for all $y \in S$, $x \in \bar{S}$ respectively,

(*iv*) For every sequence $\{y^k\} \subseteq S$, if $\lim_{k\to\infty} y^k = y^* \in \bar{S}$ then $\lim_{k\to\infty} D(y^*, y^k) = 0$,

(*v*) For every pair of sequences $\{x^k\} \subseteq \bar{S}$, $\{y^k\} \subseteq S$, if $\lim_{k\to\infty} D(x^k, y^k) = 0$, $\lim_{k\to\infty} y^k = y^* \in \bar{S}$ and $\{x^k\}$ is bounded, then $\lim_{k\to\infty} x^k = y^*$.

Note that condition (*ii*) implies that $D(x,y) \ge 0$, $D(x,y) = 0$ iff $x = y$. D, however, is not a distance, since in general it is not symmetric (it is symmetric iff f is quadratic, in which case Bregman's algorithm reduces to Hildreth's quadratic programming method [22,23]).

By analogy with the case in which D is the Euclidean distance, define the <u>D-projection</u>, or <u>Bregman projection</u> P_c of a point $y \in S$ onto a closed convex set $C \subseteq \mathbb{R}^n$ such that $C \cap \bar{S} \neq \emptyset$, as:

$$P_c(y) = \arg\min_{x \in C \cap \bar{S}} D(x,y) \ . \tag{6}$$

It is shown in [11] that conditions (i)-(v) guarantee that P_c is well defined. In the case of <u>equality constraints</u> and <u>no relaxation</u> Bregman's algorithm consists of successive D-projections onto the hyperplanes defined by the constraints. More precisely, assume that a^i are the rows of A, and let $H_i = \left\{ x \mid \langle a^i, x \rangle = b_i \right\}$. Consider a <u>control sequence</u> $\{i(k)\}$, $k=1,2,\ldots$, with $i(k) \in \{1,2,\ldots,m\}$. $i(k)$ is the index of the row used at iteration k. $i(k)$ is <u>almost cyclic</u> if there exists $M \geq m$ such that for all i $(1 \leq i \leq m)$ and all $k \geq 0$ there exists ℓ such that $k \leq \ell \leq k+M$ and $i(\ell)=i$ (i.e., every constraint is used at least once in every M consecutive iterations). When M=m the control sequence is <u>cyclic</u>.

Bregman's algorithm in this case is:

$$\begin{cases} x^0 \in S, \\ x^{k+1} = P_{H_{i(k)}}(x^k) \end{cases} \tag{7}$$

where $\{i(k)\}$ is an almost cyclic control sequence.

By (6), $P_{H_i}(y)\in\bar{S}$ for all $y\in S$ $(1\le i\le m)$, but since the domain of P_H is S rather than $\bar{S}$, in order for (7) to make sense we introduce the following definition: f is <u>zone consistent</u> with respect to a hyperplane H if $H\cap S\ne\emptyset$ and $P_H(y)\in S$ for all $y\in S$. We then add another hypothesis (on both the objective function and the constraints):

(ω_i) f is zone consistent with respect to H_i $(1\le i\le m)$.

In the case of inequalities, it is reasonable to expect that dual variables (i.e., Kuhn-Tucker multipliers) be required. In this case the algorithm generates two sequences: $\{x^k\}\subseteq\mathbb{R}^n$, $\{z^k\}\subseteq\mathbb{R}^m$ and x^{k+1} is the D-projection of x^k onto a hyperplane $H'(k)$ parallel to $H_{i(k)}$ and lying between x^k and $H_{i(k)}$, whose location is determined by the current value of the component of the dual variable associated with constraint $i(k)$, i.e., $z^k_{i(k)}$.

In order to give a formal definition, it is convenient to give a more explicit representation of P_H for a hyperplane $H=\{x\,|\,\langle a,x\rangle=b\}$. It is shown in [11] that $P_H(y)$ is the unique solution x of:

$$\begin{cases} \nabla f(x) = \nabla f(y) + sa\ , & (8)\\ \langle a,x\rangle = b\ , & (9)\\ x\in\bar{S}\ . & (10)\end{cases}$$

These are the Kuhn-Tucker conditions for problem (6) when C=H, sufficient by convexity of D in its first variable. $s\in\mathbb{R}$ is the

Lagrange multiplier associated with the constraint $x \in H$, it is uniquely determined by (8)-(10) given H and y, and it will be denoted by $\pi_H(y)$.

Since in this case x^k is projected onto a hyperplane which is not $H_{i(k)}$, we need something more than zone consistency. So we say that f is <u>strongly zone consistent</u> with respect to a hyperplane H iff, for all $y \in S$, $P_{H'}(y) \in S$ for all hyperplanes H' parallel to H lying between y and H, and replace condition (vi) by:

(vi') f is strongly zone consistent with respect to every H_i, $1 \le i \le m$

The <u>unrelaxed Bregman algorithm</u> for the problem

$$\begin{cases} \min f(x) \\ \text{s.t. } Ax \le b\,, \end{cases} \tag{11}$$

where f, A, and b satisfy conditions (i)-(v) and (vi'), is given by:

$i)$ $z^o \in \mathbb{R}^m_{\ge}$ is artibtrary , (12)

and $x^o \in S$ such that $\nabla f(x^o) = -A^T z^o$, (13)

where A^T is the transposed of the Matrix A .

$ii)$ Given x^k and z^k, x^{k+1} and z^{k+1} are uniquely determined by:

$$\nabla f(x^{k+1}) = \nabla f(x^k) + c_k a^{i(k)}, \tag{14}$$

$$z^{k+1} = z^k - c_k e^{i(k)} . \tag{15}$$

with

$$c_k = \min\left\{z^k_{i(k)}, \beta_k\right\}, \tag{16}$$

$$\beta_k = \pi_{H_{i(k)}}(x^k), \tag{17}$$

where $\{i(k)\}$ is an almost cyclic control sequence and $e^i_j=1$ if $i=j$, 0 otherwise.

Finally we introduce under relaxation. Consider a sequence of <u>relaxation parameters</u> $\{\alpha_k\}$ where $\epsilon\le\alpha_k\le 1$ for some $\epsilon>0$ and all $k\ge 0$. The <u>underrelaxed Bregman algorithm</u>for problem (11), where f, A, and b satisfy conditions (i)-(v) and (iv'), is given by (12)-(16) plus:

$$\beta_k = \pi_{\tilde{H}(k)}(x^k), \tag{17'}$$

where

$$\tilde{H}(k) = \left\{x \mid \langle a^{i(k)}, x\rangle = \alpha_k b_{i(k)} + (1-\alpha_k)\langle a^{i(k)}, x^k\rangle\right\}, \tag{18}$$

and where $\{i(k)\}$ is an almost cyclic control sequence and the sequence $\{\alpha_k\}$ satisfies:

$$0<\epsilon\le\alpha_k\le 1 . \tag{19}$$

Any sequence $\{x^k\}$ generated by the underrelaxed Bregman algorithm (12-16), (17'), (18)-(19) converges to the solution of problem (11), as proved in [15].

Note that the implicit definition of x^{k+1} in (14)-(17) makes it necessary to solve two systems of non-linear equations at each iteration. First the system of n+1 equations:

$$\begin{cases} \nabla f(x)=\nabla f(x^k)+sa^{i(k)} & (20) \\ \langle a^{i(k)},x\rangle=\alpha_k b_{i(k)}+(1-\alpha_k)\langle a^{i(k)},x^k\rangle & (21) \end{cases}$$

has to be solved in the n+1 unknowns $x_1,\ldots,x_n,s$, and β_k is the value of s. Next, if $c_k\neq\beta_k$, the system of n equations in the n unknowns $x_1,\ldots,x_n$:

$$\nabla f(x)=\nabla f(x^k)+c_k a^{i(k)} \tag{22}$$

is solved and its solution is x^{k+1}. The need to repeatedly solve these two systems numerically may strongly reduce the practicality of Bregman's algorithm. In many cases, however, the function f is such that the system (8) can be solved explicity for x in terms of y,a and s, i.e., there exists a function of 2n+1 variables h(y,a,s) with values in $\mathbb{R}^n$, with an explicit analytic expression, such that x=h(y,a,s) is the solution of (8). In such a case, the solution of (8)-(10) reduces to solving one non-linear equation. We define $g:\mathbb{R}\to\mathbb{R}$ as

$$g(s)=\langle a,h(y,a,s)\rangle \tag{23}$$

and (9) becomes

$$g(s)=b\ . \tag{24}$$

This equation might have more than one solution but if f is a Bregman function, zone consistent with respect to the hyperplane $\{x|\langle a,x\rangle=b\}$, then, because of the uniqueness of the D-projection, only one of them, say s^*, will be such that $h(y,a,s^*)\in\bar{S}$. This is the case for the "x log x-entropy" function, and also, as we show next, for Burg's entropy function.

3. UNDERRELAXED BREGMAN'S ALGORITHM FOR BURG'S ENTROPY SUBJECT TO LINEAR INEQUALITY CONSTRAINTS

Before discussing the applicability of Bregman's algorithm to Burg's entropy function (i.e., the validity of (i)-(vi')) we specialize it for this case.

The natural zone for $-B(x)$, as defined by (1), is $S=\mathbb{R}^n_>$, the interior or $\mathbb{R}^n_{\geq}$. It follows from (1) and (8) that the appropriate function h is given by:

$$h_j(y,a,s) = \frac{y_j}{1-a_j y_j s}\,, \tag{25}$$

so that g is given by:

$$g(s) = \sum_{j=1}^{n} \frac{a_j y_j}{1-a_j y_j s}\,. \tag{26}$$

The equation $g(s)=b$ may have many solutions. In order to ensure that $P_H(y)\in S$, i.e., that $h(y,a,s)>0$, the solution s^* has to be in the interval:

$$r<s^*<t \tag{27}$$

where

$$r=\max\{1/a_j y_j \mid 1\leq j\leq n,\ a_j<0\}\,, \tag{28}$$

and if $a_j\geq 0$ for all j, set $r=-\infty$, and

$$t=\min\{1/a_j y_j \mid 1\leq j\leq n,\ a_j>0\} \tag{29}$$

and if $a_j\leq 0$ for all j, set $t=+\infty$.

A direct proof that, for $y>0$, there exists only one s satisfying (27) can be found in [12, Proposition 2]. Combining (25)-(29) with the algorithm (12)-(16), (17'), (18) we get:

<u>Algorithm 1.</u> <u>Underrelaxed Bregman's algorithm for inequality constrained Burg's entropy maximization:</u>

Let $\{i(k)\}$ be an almost cyclic control sequence, $\{\alpha_k\}$ a sequence of relaxation parameters satisfying, for some ε:

$$0<\varepsilon\le\alpha_k\le 1 . \tag{30}$$

Let $e^i\in\mathbb{R}^m$ be defined as $e^i_i=1$, $e^i_j=0$ for $i\neq j$.

<u>Initialization</u>: $z^0\in\mathbb{R}^m_>$ arbitrary, and (31)

$$x^0_j=\frac{1}{(A^T z^0)_j} , \quad 1\le j\le n . \tag{32}$$

<u>Iterative Step</u>: Given x^k, z^k:

(i) Let β_k be the unique solution s of:

$$\sum_{j=1}^{n}\frac{a_j^{i(k)}x_j^k}{1-a_j^{i(k)}x_j^k s}=\alpha_k b_{i(k)}+(1-\alpha_k)\langle a^{i(k)},x^k\rangle , \tag{33}$$

for which

$$r_k<\beta_k<t_k \tag{34}$$

where

$$r_k=\max\left\{1/a_j^{i(k)}x_j^k \mid 1\le j\le n,\ a_j^{i(k)}<0\right\} \tag{35}$$

and if $a_j^{i(k)}\ge 0$ for all j, then $r_k=-\infty$, and

$$t_k = \min\left\{1/a_j^{i(k)} x_j^k \mid 1\le j\le n,\ a_j^{i(k)}>0\right\} \tag{36}$$

and if $a_j^{i(k)}\le 0$ for all j, then $t_k=+\infty$.

(*ii*)
$$c_k = \min\left\{z^k_{i(k)},\ \beta_k\right\}. \tag{37}$$

(*iii*)
$$x_j^{k+1} = \frac{x_j^k}{1-a_j^{i(k)} x_j^k c_k}, \quad 1\le j\le n. \tag{38}$$

(*iv*)
$$z^{k+1} = z^k - c_k e^{i(k)}. \tag{39}$$

4. CONVERGENCE ANALYSIS FOR ALGORITHM 1

A complete proof for the underrelaxed Bregman's algorithm for a general Bregman objective function was given in [15]. If the function $[-B(x)]$, as defined by (1), were such that conditions (*i*)-(*v*) hold (and A and b such that (*vi'*) holds), then the convergence results of [15] would be directly applicable.

This is, however, not the case. The function $f(x)=-B(x)=-\sum_{j=1}^{n}\log x_j$ is not continuous on the boundary of the nonnegative orthant, and in fact
$$D(x,y) = \sum_{j=1}^{n}\left[\log\left(\frac{y_j}{x_j}\right) - \left(\frac{x_j}{y_j}\right)\right]+n$$

is defined only on $S \times S$, rather than $\bar{S} \times S$, so that conditions (ii), (iv) and (v) do not hold. However, the convergence analysis of [15] can be modified, as we show next. First, observe that for $f(x)=-B(x)$, (i)-(v) hold with S substituting for $\bar{S}$ (i.e., in the interior of the positive orthant). (i) and (ii) are immediate. (iii) is a consequence of the fact that

$$\lim_{\|x\| \to \infty} D(x,y)=+\infty \ , \quad \lim_{\|y\| \to \infty} D(x,y)=+\infty \ , \tag{40}$$

for all y, respectively all x, in $\mathbb{R}^n_>$.

Also, for every j, $j=1,2,\ldots,n$,

$$\lim_{x_j \to 0^+} D(x,y)=+\infty \ , \quad \lim_{y_j \to 0^+} D(x,y)=+\infty \ , \tag{41}$$

for all y, respectively all x, in $\mathbb{R}^n_>$.

Regarding (iv) and (v), when $y^* \in S$ and $\{x^k\} \subseteq S$, they readily follow from the continuity of D in both variables (consequence of (i) and (4)) and the already mentioned fact that, for $x,y \in S$, $D(x,y)=0$ iff $x=y$.

Next observe that under strong zone consistency the whole sequence $\{x^k\}$ is contained in S, because x^{k+1} is always the D-projection of x^k onto a hyperplane parallel to one of the H_i's lying between it and x^k (the validity of the strong zone consistency hypothesis for this case is considered below).

This means that statements regarding a finite subset of the sequences $\{x^k\}$ and $\{z^k\}$ remain valid if (i)-(v) hold in S rather than $\bar{S}$. Two such relevant statements, refered to the sequences $\{x^k\}$, $\{z^k\}$ defined by (12)-(16), (17'), (18), are:

(a) The sequence $\left\{f(x^k)+\langle z^k, Ax^k-b\rangle\right\}$ is decreasing, and

(b) For all $x\in S$ such that $Ax\le b$, $D(x,x^k)\le f(x)-f(x^k)-\langle z^k, Ax^k-b\rangle$.

(a) is proved in [15, Corollary 4.1] and (b) in [15, Proposition 4.6]. An immediate consequence of (a) and (b) is that for all x feasible for (11):

$$D(x,x^k) \le f(x)-f(x^o) - \langle z^o, Ax^o-b\rangle \ . \tag{42}$$

Calling the right hand side of (42) γ^*, it follows from (5) that, when the system $Ax\le b$ is feasible, then for all k and all feasible x:

$$x^k\in L_2(x,\gamma^*) \ . \tag{43}$$

But (40) and (41) imply that $L_2(x,\gamma^*)$ is a bounded and closed subset of S, which contains the whole sequence $\{x^k\}$, when $f(x)=-B(x)$. Hence all limit points of $\{x^k\}$ are in S.

In fact, hypotheses on the behavior of f and D on the boundary of S are introduced, because for a general Bregman fuction, although the sequence $\{x^k\}$ is contained in S, its limit points may be on the boundry (such is the case for instance, with the "x log x-entropy" function). We have just seen that this cannot happen with Burg's entropy. Once we have established that all limit points of $\{x^k\}$ are in S, the remainder of the convergence analysis of [15] is applicable with hypotheses (ii), (iii), (v) holding in S (rather than $\bar{S}$). We still need to check (vi'), which relates in addition to f, also to A and b.

Let
$$F = \{x>0 \mid Ax\leq b\}, \quad (44)$$
$$N(A) = \{x\in\mathbb{R}^n \mid Ax\leq 0\}. \quad (45)$$

If follows from [11, Lemma 1] that the following conditions insure hypothesis (vi').

$$a^i \neq 0 \quad , 1\leq i\leq m , \quad (46)$$
$$F \neq \emptyset , \quad (47)$$
$$N(A)\cap \mathbb{R}^n_+ = \{0\}. \quad (48)$$

Condition (46) is required in order to have hyperplanes defined by the constraints; (47) is just feasibility of the constraint set, an obvious condition for convergence of Bregman's algorithm in general, and (48) insures that all relevant hyperplanes intersect S. This guarantees existence of the D-projections, which cannot be in the boundary of S as a consquence of (41), so that (vi') holds.

The previous analysis shows that the convergence results of [15] hold indeed in this case, meaning that when A and b satisfy (46)-(48) then the sequence $\{x^k\}$, defined by (31)-(39), converges to the solution of:

$$\begin{cases} \max \sum_{j=1}^{n} \log x_j \\ \text{s.t.} \quad Ax\leq b,\ x>0 . \end{cases} \quad (49)$$

This analysis can be extended to other Bregman functions which fail to satisfy (i)-(iii') because of singularities in the boundary of S. Basically, for such functions an additional hypothesis is required, namely, that the sets $L_1(y,\gamma)$, $L_2(x,\gamma)$ be bounded away from such singularities. This was worked out in [7].

5. THE HYBRID ALGORITHM

Numerical solution of (33) at each iteration of Bregman's algorithm is an undesirable feature. The algorithm defined in this section avoids such a situation by dynamically choosing a sequence of relaxation parameters $\{\alpha_k\}$ (where α_k depends on A,b and also on the current iterate x^k) so that (33) has an explicit solution. In general, assume that we are dealing with a function f such that a closed formula h(y,a,s) gives the solution x of (8). Then, at iteration k we define, as in (23):

$$g_k(s) = \langle a^{i(k)}, h(x^k, a^{i(k)}, s)\rangle , \tag{50}$$

and β_k is the solution s of:

$$g_k(s) = \alpha_k b_{i(k)} + (1-\alpha_k)\langle a^{i(k)}, x^k\rangle . \tag{51}$$

Condition (19) means that we are looking for s such that $g_k(s)$ is in the segment between $\langle a^{i(k)}, x^k\rangle$ and $b_{i(k)}$. Since $g_k(0)=\langle a^{i(k)}, x^k\rangle$, when g_k is continuous and monotonous near 0, we are seeking some s between 0 and the solution s^* of $g_k(s)=b_{i(k)}$. From (19) and (51) we need an s such that for some $\varepsilon>0$ and all k:

$$\epsilon \le \frac{g_k(s) - \langle a^{i(k)}, x^k \rangle}{b_{i(k)} - \langle a^{i(k)}, x^k \rangle} \le 1 \quad (52)$$

Therefore, if we are able to present a formula

$$s = \varphi(a, y, b) \quad (53)$$

such that for

$$s = s_k = \varphi(a^{i(k)}, x^k, b_{i(k)}) \quad (54)$$

condition (51) is satisfied, we may take $\beta_k = s_k$ and such a β_k will be the one given by (17') when the coefficients α_k used in (18) are defined as:

$$\alpha_k = \frac{g_k(\varphi(a^{i(k)}, x^k, b_{i(k)})) - \langle a^{i(k)}, x^k \rangle}{b_{i(k)} - \langle a^{i(k)}, x^k \rangle} \quad (55)$$

where (51) guarantees that (19) holds.

Note that if $b_{i(k)} = \langle a^{i(k)}, x^k \rangle$ then the solution of (51) is $s=0$, and $x^{k+1} = x^k$ (i.e., if $x^k \in H_{i(k)}$, then $\tilde{H}(k) = H_{i(k)}$ and the D-projection of x^k is just x^k). In the case of Burg's entropy, such a function φ is available, if we impose the following additional condition on A and b (which, by the way, implies (48)):

$$a^i_j b_i \ge 0 , \quad 1 \le i \le m, \quad 1 \le j \le n . \quad (56)$$

The appropriate φ turns out to be:

$$\varphi(a, y, b) = \begin{cases} \lambda(1 - \frac{\langle a, y \rangle}{b})t , & \text{if } b > 0 , \\ \lambda(1 - \frac{\langle a, y \rangle}{b})r , & \text{if } b < 0 , \end{cases} \quad (57)$$

where r,t are given by (28), (29) and $0<\lambda<1$.

<u>Algorithm 2. Hybrid algorithm for inequality constrained Burg's entropy maximization</u>.

Let $\{i(k)\}$ be an almost cyclic control sequence, let $e^i \in \mathbb{R}^m$ be defined as $e^i_i=1$, $e^i_j=0$ for $i\neq j$. Consider a sequence $\{\lambda_k\}$ such that for some η

$$0\le\eta\le\lambda_k\le1 \ , \tag{58}$$

and assume that A and b satisfy (46), (47), and (56).

The algorithm is the same as (31)-(39) except that substep $i)$ of the iterative step is replaced by:

$$\beta_k = \lambda_k\left(1-\frac{\langle a^{i(k)},x^k\rangle}{b_{i(k)}}\right)\theta_k \tag{59}$$

where

$$\theta_k = \begin{cases} t_k \ , & \text{if } b_{i(k)}>0 \ , \\ r_k \ , & \text{if } b_{i(k)}<0 \ , \end{cases} \tag{60}$$

with r_k, t_k defined as in (35), (36).

6. CONVERGENCE ANALYSIS FOR ALGORITHM 2

Convergence of Algorithm 2 to the solution of problem (49) is an immediate consequence of the analysis of Section 4 once we establish that (19) holds for α_k defined as in (55), where g_k is given (as in

(26)) by

$$g_k(s) = \sum_{j=1}^{n} \frac{a_j^{i(k)} x_j^k}{1-a_j^{i(k)} x_j^k s} \tag{61}$$

and σ is given by (57).

<u>Proposition 1.</u> There exists $\varepsilon>0$ such that α_k, as defined by (55), satisfies

$$\varepsilon\le\alpha_k\le 1 . \tag{62}$$

<u>Proof.</u>

We consider in detail the case $b_{i(k)}>0$. From (55), (57) and (61):

$$\alpha_k = \sum_{j=1}^{n} \frac{a_j^{i(k)} x_j^k}{\langle a^{i(k)}, x^k\rangle + \sigma_{jk}} \tag{63}$$

where

$$\sigma_{jk} = -b_{i(k)}\left(1- \frac{1}{a_j^{i(k)} x_j^k t_k \lambda_k}\right) . \tag{64}$$

Since $b_{i(k)}>0$, (56) implies

$$a_j^{i(k)}\ge 0 \ , \quad 1\le j\le n . \tag{65}$$

From (58) and (36) we get:

$$\frac{1}{a_j^{i(k)} x_j^k t_k \lambda_k} \ge 1 , \tag{66}$$

thus,

$$\sigma_{jk} \ge 0 \ , \ 1\le j\le n, \ k\ge 0 , \tag{67}$$

which proves, in view of (63), that

$$\alpha_k \le 1 . \tag{68}$$

Also, by strong zone consistency $x^k_j > 0$ for all j, so that (46), (63),(65) and (67) imply:

$$0 < \alpha_k . \tag{69}$$

We have already proved that the hybrid algorithm is a specific realization of Algorithm 1 with relaxation parameters α_k satisfying:

$$0 < \alpha_k \le 1 . \tag{70}$$

The existence of ϵ as in (62) is not required in the convergence analysis of Bregman's algorithm [15] for establishing (a) and (b) of Section 4, which imply (43). In fact, (43) holds if (70) is satisfied. So we may assume that $x^k \in L_2(x, \gamma^*)$ for some $x \in F$. Since for Burg's entropy, $L_2(x, \gamma^*)$ is a compact subset of $\mathbb{R}^n_>$ (see Section 4) there exist $q, Q \in \mathbb{R}$ such that

$$0 < q \le x^k_j \le Q , \quad 1 \le j \le n, \quad k \ge 0 . \tag{71}$$

Define,

$$p = \min\left\{ |a^i_j| \mid 1 \le i \le m , 1 \le j \le n , a^i_j \ne 0 \right\} , \tag{72}$$

$$P = \max\left\{ |a^i_j| \mid 1 \le i \le m , 1 \le j \le n \right\} \tag{73}$$

and

$$\delta = \max\left\{ |b_i| \mid 1 \le i \le m \right\} . \tag{74}$$

Considering only summands of (63) for which $a^{i(k)}_j \ne 0$ and using

(29) and (58) we conclude that:

$$\left| \frac{1}{a_j^{i(k)} x_j^k t_k \lambda_k} \right| \le \frac{PQ}{\eta pq} \tag{75}$$

because

$$\frac{1}{t_k} = \max\left\{ a_j^{i(k)} x_j^k \mid 1 \le j \le n \ , \ a_j^{i(k)} > 0 \right\} . \tag{76}$$

From (64), (74) and (75) we get

$$\sigma_{jk} \le \theta \left(1 + \frac{PQ}{\eta pq} \right) \equiv \Omega \ . \tag{77}$$

Also,

$$pq \le |\langle a^{i(k)}, x^k \rangle| \le nPQ \tag{78}$$

and, therefore, from (63) and in view of (65), we obtain finally that, for all $k \ge 0$,

$$\alpha_k \ge \frac{pq}{\eta PQ + \Omega} . \tag{79}$$

Selecting ε as the right hand side of (79) completes the proof. The case $b_{i(k)} < 0$ is handled in a similar way.

□

This proposition means that Algorithm 2 is a particular realization of Algorithm 1. Thus, the convergence analysis of Section 4 implies that the sequence $\{x^k\}$ generated by Algorithm 2 converges to the solution of problem (49).

7. THE CASE OF EQUALITY CONSTRAINTS

If the problem to be solved is:

$$\max \sum_{j=1}^{n} \log x_j \tag{80}$$
$$\text{s.t. } Ax=b,$$
$$x>0,$$

then Algorithms 1 and 2 can be applied after transforming each equation into two inequalities. A more efficient approach is to modify Algorithms 1 and 2 by changing the initialization step to:

$$x^o > 0 , \tag{81}$$

suppressing substeps *ii*) and *iv*) of the iterative step and substituting β_k for c_k in (38). As discussed in [15] the convergence analysis performed for the inequality case remains valid. The sequences $\{x^k\}$ generates by these two algorithms (which could be called Algorithms 1' and 2') are not the same as those generated by Algorithms 1 and 2, respectively, after converting each equation into two inequalities. In the case of Algorithm 1 and 1' the sequences $\{x^k\}$ are indeed the same in the unrelaxed case, i.e., when $\alpha_k=1$ for all k. Algorithm 1' is just an underrelaxed version (with almost cyclic, rather than cyclic, control) of the algorithm proposed in [10] and analyzed in [12]. Algorithm 2' is new.

ACKNOWLEDGEMENTS

Part of this work was done during a visit of Y. Censor to the Instituto de Matemática Pura e Aplicada (IMPA) in Rio de Janeiro, Brazil. We thank Professor Lindolpho de Carvalho Dias, Director of IMPA, for making this visit possible. The research of Y. Censor has been also supported by NIH grant HL-28438 while visiting the Medical Image Processing Group (MIPG) at the Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, PA.

REFERENCES

[1] M. Ben-Basat, f-entropies, probability of error, and feature selection, Information and Control 39 (1978), 227-242.

[2] J.P. Burg, Maximum entropy spectral analysis, in: Proceedings of the 37th Annual Meeting of the Society of Exploration Geophisicists, Oklahoma City, Oklahoma (1967).

[3] L.M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, U.S.S.R. Computational Mathematics and Mathematical Physics 7(3) (1967), 200-217.

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

[5] Y. Censor, Finite series-expansion reconstruction methods, Proceedings of the IEEE 71(1983), 409-419.

[6] Y. Censor, A.R. De Pierro, T. Elfving, G.T. Herman and A.N. Iusem, On iterative methods for linearly constrained entropy maximization, in Numerical Analysis and Mathematical Modelling (A. Wakulicz, Editor) Banach Center Publications, vol. XXIV. Warsaw, Poland. (To appear).

[7] Y. Censor, A.R. De Pierro and A.N. Iusem, On maximiation of entropies and a generalization of Bregman's method for convex programming. Technical Report MIPG 113, Medical Image Processing Group. Univ. of Pennsylvania, October 1986.

[8] Y. Censor, T. Elfving and G.T. Herman, Methods for entropy maximization with applications in image processing, in: Pro-

ceedings of the Third Scandinavian Conference on Image Analysis (P. Johansen and P.W. Becker, Editors), Chartwell- Bratt, Ltd., Lund, Sweden (1983), 296-300.

[9] Y. Censor, T. Elfving and G.T. Herman, Special-purpose algorithms for linearly constrained entropy maximization, Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems, (C.R. Smith and G.J. Erickson, Editors), D. Reidel Pub. Co., Dordrecht, Holland (1977), 241-254.

[10] Y. Censor, A.V. Lakshminarayanan and A. Lent, Relaxational methods for large-scale entropy optimization problems, with applications in image reconstruction, in: Information Linkage Between Applied Mathematics and Industry (P.C.C. Wang et al., Editors), Academic Press, N.Y. (1979), 539-546.

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

[12] Y. Censor and A. Lent, Optimization of "log x"-entropy over linear equality constriants, SIAM Journal on Control and Optimization, 25 (1987), 921-933.

[13] L.R. D'Addario, Maximum entropy imaging: theory and philosophy, in: Image Analysis and Evaluation, (R. Shaw, Editor), Society of Photographic Scientists and Engineers, (SPSE), Washington, D.C. (1977), 221-225.

[14] J.N. Darroch and D. Ratcliff, Generalized iterative scaling for log-linear models, The Annals of Mathematical Statistics 43 (1972), 1470-1480.

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

[16] T. Elfving, On some methods for entropy maximization and matrix scaling, Linear Algebra and Its Applications, 34 (1980), 321-339.

[17] S. Erlander, Entropy in linear programs, Mathematical Programming 21 (1981), 137-151.

[18] B.R. Frieden, Image enhancement and restoration, in: Picture Processing and Digital Filtering, (T.S. Huang, Editor), Vol. 6 in the series Topics in Applied Physics, Springer-Verlag, Berlin, Heidelberg, New York (1975), 177-248.

[19] B.R. Frieden, Statistical models for the image restoration problem, Computer Graphics and Image Processing 12 (1980), 40-59.

[20] B.R. Frieden, Probability, Statistical Optics, and Data Testing Springer-Verlag, Berlin, Heidelberg, New York, (1983).

[21] G.T. Herman, Image Reconstruction From Projections: The Fundamentals of C mputerized Tomography, Academic Press, New York, (1980).

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

[23] C. Hildreth, A quadratic programming procedure, Naval Research Logistic Quarterly 4 (1957), 79-85, Erratum, ibid., 361.

[24] A.K. Jain and S. Ranganath, Two dimensional spectral estimation Proceedings of the RADC (Rome Air Development Center) Workshop on Spectral Estimation, held at RADC, Griffis Air Base, Rome, N.Y., May 1978, 151-157.

[25] E.T. Jaynes, On the rationale of maximum-entropy methods, Proceedings of the IEEE 70 (1982), 939-952.

[26] R. Johnson and J.E. Shore, Which is the better entropy expression for speech processing: - S log S or log S ?, IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-32 (1984), 129-136.

[27] J.N. Kapur, Twenty-five years of maximum-entropy principle, Journal of Mathematical and Physical Sciences 17 (1983), 103-156.

[28] J.N. Kapur, A comparative assessment of various measures of entropy, Journal of Information and Optimization Sciences 4 (1983), 207-232.

[29] B. Lamond and N.P. Steward, Bregman's balancing method, Transportation Research 15B (1981), 239-248.

[30] A. Lent, A convergent algorithm for maximum entropy image restoration with a medical X-ray application, in: Image Analysis and Evaluation, (R. Shaw, Editor), Society of Photographic Scientists and Engineers, (SPSE), Washington, D.C. (1977), 249-257.

[31] S.J. Wernecke, Maximum entropy techniques for digital image reconstruction, in: Image Analysis and Evaluation (R. Shaw, Editor), Society of Photographic Scientists and Engineers

(SFSE), Washington, D.C. (1977), 238-243.

[32] D.S. Wong, Maximum likelihood, entropy maximization, and the geometric programming approaches to the calibration of trip distribution models, Transportation Research 15B (1981), 329-343.

Yair Censor
Department of Mathematics and Computer Science
University of Haifa, Haifa 31.999, Israel

Alvaro R. De Pierro
IMECC-UNICAMP
Caixa Postal 6065
13.081 Campinas, SP, Brazil

Alfredo N. Iusem
Instituto de Matemática Pura e Aplicada (IMPA)
Estrada Dona Castorina 110,
CEP 22.460 Rio de Janeiro, RJ. Brazil

RELATÓRIOS TÉCNICOS — 1989

01/89 — Uniform Approximation of Continuous Functions With Values in [0, 1] — *João B. Prolla.*

02/89 — On Some Nonlinear Iterative Relaxation Methods in Remote Sensing — *A. R. De Pierro.*

03/89 — A Parallel Iterative Method for Convex Programming with Quadratic Objective — *Alfredo N. Iusem* and *Alvaro R. De Pierro.*

04/89 — Fifth Force, Sixth Force, and all that: a Theoretical (Classical) Comment — *Erasmo Recami* and *Vilson Tonin–Zanchin.*

05/89 — An Application of Singer's Theorem to Homogeneous Polynomials — *Raymundo Alencar.*

06/89 — Summhammer's Experimental Test of the Non–Ergodic Interpretation of Quantum Mechanics — *Vincent Buonomano.*

07/89 — Privileged Reference Frames in General Relativity — *Waldyr A. Rodrigues Jr.* and *Mirian E. F. Scanavini.*

08/89 — On the Numerical Solution of Bound Constrained Optimization Problems — *Ana Friedlander* and *José Mario Martínez.*

09/89 — Dual Extremun Principles for the Heat Equation Solved by Finite Element Methods I — *Vera Lucia da Rocha Lopes* and *José Vitório Zago.*

10/89 — Local Convergence Theory of Inexact Newton Methods Based on Structured Least Chance Updates — *José Mario Martínez*

11/89 — Real Spin–Clifford Bundle and the Spinor Structure of Space–Time — *Waldyr A. Rodrigues Jr.* and *Vera L. Figueiredo.*

12/89 — A Multiplier Theorem on Weighted Orlicz Spaces — *B. Bordin* and *J. B. Garcia.*

13/89 — Dual Extremum Principles For The Heat Equation Solved By Finite Element Methods II — *Vera Lucia da Rocha Lopes* and *José Vitório Zago.*

14/89 — Dirac and Maxwell Equations in the Clifford and Spin–Clifford Bundels — *W. A. Rodrigues Jr.* and *E. Capelas de Oliveira.*

15/89 — Formal Structures, The Concepts of Covariance Invariance, Equivalent Reference Frames, and the Principle of Relativity — *W. A. Rodrigues Jr., M. E. F. Scanavini* and *L. P. de Alcantara.*

16/89 — Local Minimizers of a Quadratic Function With a Spherical Constraint — *José Mario Martínez.*

17/89 — On Pseudo–Convex Polycircular Domains In Banach Spaces — *Mário C. Matos.*

18/89 — On Circular and Special Units of an Abelian Number Field — *Trajano Nóbrega.*

19/89 — **Implementing Algorithms for Solving Sparse Nonlinear Systems of Equations** — *Márcia A. Gomes-Ruggiero, José Mario Martínez* and *Antonio Carlos Moretti.*

20/89 — **An Algorithm for Solving Nonlinear Least-Squares Problems with a New Curvilinear Search** — *José Mario Martínez* and *Rita Filomena Santos.*

21/89 — **Covariant Spinors, Algebraic Spinors, Operator Spinors and their Relationship** — *V. L. Figueiredo, E. Capelas de Oliveira* and *W. A. Rodrigues Jr.*

22/89 — **The Column-Updating Method for Solving Nonlinear Equations in Hilbert Space** — *Márcia A. Gomes-Ruggiero* and *José Mario Martínez.*

23/89 — **Automorphisms Control Systems and Observability** — *Victor Ayala Bravo*

24/89 — **Adjoint Functors Arise Everywhere** — *T. M. Viswanathan.*

25/89 — **New Sum Rules of Special Functions** — *E. Capelas de Oliveira.*

26/89 — **Minimal Realizations Under Controllability** — *Luiz A. B. San Martin* and *Victor Ayala Bravo.*