Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

P.P.B. Eggermont

Multiplicative iterative algorithms for convex programming

Department of Applied Mathematics Report AM-R8808 July

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

Multiplicative Iterative Algorithms for Convex Programming

P.P.B. Eggermont
Centre for Mathematics and Computer Science
PO Box 4079, 1009 AB Amsterdam, The Netherlands
and
Department of Mathematical Sciences
University of Delaware, Newark DE 19716, USA

We study multiplicative iterative algorithms for the minimization of a differentiable, convex function defined on the positive orthant of $\mathbb{R}^N$. If the objective function has compact level sets and has a locally Lipschitz continuous gradient then these algorithms converge to a solution of the minimization problem. Moreover, the convergence is nearly monotone in the sense of Kullback-Leibler divergence. Special cases of these algorithms have been applied in position emission tomography and are formally related to the EM algorithm for positron emission tomography

Key words: multiplicative iterative algorithms, convex optimization, Kullback-Liebler divergence, positron emission tomography, EM algorithm.
1980 Mathematics Subject Classification: 65K05, 90C30, 90C20, 92A07

1. Introduction

In this paper we study multiplicative iterative algorithms for the minimization problem

$$\begin{aligned} &\text{Minimize } l(x) \\ &\text{subject to } x \geqslant 0 \end{aligned} \tag{1.1}$$

where l is a convex, continuously differentiable function on $\mathbb{R}^N$ with compact level sets, and locally Lipschitz continuous gradient. The interest in such algorithms is sparked by the emergence of the EM algorithm for maximum likelihood estimation in positron emission tomography, Shepp and Vardi [8], and by the more or less ad hoc variation proposed by Daube-Witherspoon and Muehllehner [2]. The EM algorithm belongs to a class of algorithms which arise as the method of successive substitution for the complementarity equations of the Kuhn-Tucker conditions for a solution of (1.1), viz. x is a solution of (1.1) if and only if

$$\begin{aligned} &x \geqslant 0, \quad \nabla l(x) \geqslant 0, \\ &x_j\,[\nabla l(x)]_j = 0, \;\; j=1,2,\ldots,N, \end{aligned} \tag{1.2}$$

see, e.g., Mangasarian [5]. The resulting algorithms have the form

$$x_j^{n+1} = x_j^n\,(1 - \omega_n\,[\nabla l(x^n)]_j), \;\; j=1,2,\ldots,N, \tag{1.3}$$

where ω_n is a relaxation/steplength parameter. Unfortunately these algorithms are rather complicated to analyse, even in the special case of the EM algorithm, Vardi et al. [10]. One notable feature of the convergence proof of the EM algorithm is the predominant role played by the *Kullback-Leibler informational divergence* between two vectors in $\mathbb{R}_+^N$, which we feel is only partly explained by the fact that the negative log-likelihood function up to a constant can be written as a Kullback-Leibler divergence, Vardi et al. [10].

The alternative interpretation is to consider these multiplicative algorithms as certain approximate "proximal point methods" for (1.1), Rockafellar [7], viz. x^{n+1} is determined/interpreted as an approximate solution of the equation

Report AM-R8808
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

$$x_j + \omega_n x_j[\nabla l(x)]_j = x_j^n, \quad j=1,2,\ldots,N, \tag{1.4}$$

or equivalently, via the Kuhn-Tucker conditions, as an approximate solution of the problem

$$\text{minimize} \quad l(x) + \omega_n^{-1} d(x^n \| x) \tag{1.5}$$
$$\text{subject to} \quad x \geqslant 0$$

where $d(x \| y)$ is the Kullback-Leibler divergence, see [4] or [10],

$$d(x \| y) = \sum_{j=1}^{N} x_j \log(x_j/y_j) + y_j - x_j. \tag{1.6}$$

Since $d(x \| y) \geqslant 0$ always, and $d(x \| y) \approx \|x-y\|_2^2$ for $x \geqslant 0, y > 0$ and $(1 - x_j/y_j)$ close to zero for all j, the problem (1.5) is indeed closely related to the standard proximal point algorithms, ROCKAFELLAR [7]. We will not exploit this connection.

Besides the "exact" algorithm

$$x_j^{n+1} (1 + \omega_n [\nabla l(x^{n+1}) = x_j^n, \quad j=1,2,\ldots,N, \tag{1.7}$$

we can now state the more practical approximate method

$$x_j^{n+1} = x_j^n / (1 + \omega_n [\nabla l(x^n)]_j), \quad j=1,2,\ldots,N. \tag{1.8}$$

We refer to these two algorithms as the implicit and explicit algorithm, respectively. After appropriate scaling, the ISRA method of DAUBE-WITHERSPOON and MUEHLLEHNER [2] is of the form (1.8).

The implicit algorithm (1.6) has a decided theoretical advantage, but is not very practical as is, whereas (1.7) has the same *computational complexity* as algorithm (1.2) and is still easy to analyze.

The key ingredients in the convergence proof of the implicit algorithm (1.6) are two kinds of monotonicity. On the one hand we have that

$$l(x^n) > l(x^{n+1}) \quad \text{unless } x^n = x^{n+1}, \tag{1.9}$$

as well as the unexpected

$$d(x^* \| x^n) \geqslant d(x^* \| x^{n+1}), \tag{1.10}$$

where x^* is any accumulation point of $\{x^n\}_n$, thus showing that there is at most one accumulation point. The existence of such an accumulation point follows from the boundedness of $\{x^n\}_n$ (from (1.9) and the compact level sets assumption on l), and is easily shown to be a solution of our minimization problem (1.1). The analysis of the explicit algorithm (1.8) is only slightly more complicated, but we need to make a suitable choice for ω_n. Apart from establishing the crucial inequalities (1.9)-(1.10) these convergence proofs are virtually identical with the one given by VARDI et al. [10] Appendix, for the convergence of the EM algorithm.

On the theoretical side this study sheds new light on the emergence of the Kullback-Leibler divergence in the theory of multiplicative iterative algorithms. There is a modest contribution on the practical side since a special case of algorithm (1.8) is the image space reconstruction algorithm of DAUBE-WITHERSPOON and MUEHLLEHNER [2] for emission tomography: the algorithm converges regardless of whether the solution is unique or not. In the case of uniqueness the convergence had been shown by DE PIERRO [3]. It should also be noted that an inequality similar to (1.10) for the original multiplicative iterative algorithm (1.3) does *not* seem to be available, except for the EM algorithm. See VARDI et al. [10], Appendix, for a complicated proof.

Finally, since every scientific endeavor seems to need an acronym we propose to describe our algorithms as MIRA: multiplicative iterative relaxation algorithms.

2. ASSUMPTIONS

In this section we state the precise assumptions on the objective function and make a choice for the relaxation parameters in the explicit algorithm.

We make three assumptions about l.

2.1 The objective function $l(x)$ is continuously differentiable and convex on $\mathbb{R}^n_+$, so for all $x,y \in \mathbb{R}^n_+$,

$$<\nabla l(x) - \nabla l(y), x - y> \geq 0, \tag{2.1}$$

cf. MANGASARIAN [5]. Here $<\cdot,\cdot>$ is the usual innerproduct on $\mathbb{R}^N$.

2.2 The level sets of l are compact, i.e. for every $y \in \mathbb{R}^N_+$ the set

$$\{ x \in \mathbb{R}^N_+ : l(x) \leq l(y)\} \text{ is compact.} \tag{2.2}$$

2.3 The objective function $l(x)$ has a locally Lipschitz continuous gradient, i.e. for every compact subset C of $\mathbb{R}^N_+$ there exists a constant $L < \infty$ such that for all $x,y \in C$

$$\|\nabla l(x) - \nabla l(y)\|_2 \leq L\,\|x - y\|_2 . \tag{2.3}$$

We choose the relaxation parameter ω_n in algorithm (1.8) as $\omega_n = \omega(x^n)$ where

$$\omega(x) = (\max\{M, 2\|\nabla l(x)\|_\infty, 2\|Lx - \tfrac{1}{2}\nabla l(x)\|_\infty\})^{-1}. \tag{2.4}$$

in which M is an arbitrary, fixed constant, and L is the Lipschitz constant of $\nabla l(x)$ for the line segment

$$\{x - t\,D(x)\nabla l(x) : 0 \leq t \leq (2\|\nabla l(x)\|_\infty)^{-1}\}. \tag{2.5}$$

Here $D(x)$ is the diagonal matrix with diagonal elements $x_1, x_2, \ldots, x_N$. For algorithm (1.7) any ω_n will do, as long as $\inf \omega_n > 0$. The existence of $x^{n+1} \geq 0$ as "defined" by (1.6) follows from its interpretation as the exact, unique solution of (1.4): note that $d(x^n \| x)$ is convex in x and tends to infinity as x does so.

It is obvious that we need that $1 + \omega_n [\nabla l(x)]_j > 0$ for all j, so $\omega(x) \leq (2\|\nabla l(x)\|_\infty)^{-1}$ is a reasonable choice. The modification (2.4) is motivated by the desire that $l(x^n) > l(x^{n+1})$, see Lemma 3.1.

We comment briefly on the above assumptions. The assumptions 2.1 and 2.2 are pretty much essential. The assumption 2.3 justifies the particular choice of $\omega(x)$ for which we can prove that $\{l(x^n)\}_n$ is decreasing. Any alternative assumption 2.3 and choice of $\omega(x)$ for which this holds true would be fine. This observation might prove valuable when generalizing the multiplicative iterative algorithms to nondifferentiable convex objective functions for which then subgradients must be employed.

3. THE DOUBLE MONOTONICITY OF THE IMPLICIT ALGORITHM

In this section we give the deceptively simple proofs of the monotonicity relations (1.9) and (1.10).

We define the mapping $H: \mathbb{R}^N_+ \to \mathbb{R}^N_+$ as follows. For $x \geq 0$ we let y be the solution of the equation

$$y + \omega(x)\,D(y)\,\nabla l(y) = x \tag{3.1}$$

and define $H(x)$ as $H(x) = y$.

LEMMA 3.1. *For any $x \in \mathbb{R}^N_+$ we have $l(x) > l(H(x))$ unless $x = H(x)$.*

PROOF. Let $y = H(x)$. By convexity we have

$$l(x) - l(y) \geq <\nabla l(y), x - y> = \omega(x) <\nabla l(y), D(y)\nabla l(y)>.$$

Since $y \geq 0$ this shows that $l(x) - l(y) > 0$ unless $D(y)\nabla l(y) = 0$ in which case $x = y = H(x)$. □

We state the second monotonicity result in slightly different form. Let m be any number with

$$\inf\{l(x): x \geqslant 0\} \leqslant m \leqslant \inf\{l(x^n): \Omega \in \mathbb{N}\} \tag{3.2}$$

and let

$$\Omega = \{x^* \in \mathbb{R}_+^N : l(x^*) \leqslant m\}. \tag{3.3}$$

LEMMA 3.2. *Let $x^* \in \Omega$. For any $x^0 > 0$ we have $d(x^* \| x^n) \geqslant d(x^* \| x^{n+1})$ for all n.*

PROOF Since $x^0 > 0$ so are all x^n and thus $d(x^* \| x^n)$ is well defined for all n. Now

$$d(x^* \| x^n) - d(x^* \| x^{n+1}) = \sum_j x_j^* \log(x_j^{n+1}/x_j^n) + x_j^n - x_j^{n+1}$$

$$= \sum_j x_j^* \log(1 + \omega_n [\nabla l(x^{n+1})]_j)^{-1} + \omega(x) x_j^{n+1} [\nabla l(x^{n+1})]_j.$$

Since $\log(1+t)^{-1} = -\log(1+t) \geqslant -t$ we thus get that the above expression is greater than or equal to

$$\omega(x) < \nabla l(x^{n+1}), x^{n+1} - x^* > \geqslant \omega(x) [l(x^{n+1}) - l(x^*)],$$

with the last inequality due to convexity. By the choice of $x^* \in \Omega$ the last expression is nonnegative. □

The convergence of the implicit algorithm is now easily shown along the general outline of the convergence proof of the EM algorithm, VARDI et al. [10], Appendix. Since it it essentially the same as for the explicit algorithm, we omit it here.

It is interesting to note that the above lemmas have analogues in the continuous case

$$\begin{cases} \dfrac{dx}{dt} = -\omega(x) D(x) \nabla l(x), \quad t > 0 \\ x(0) > 0 \end{cases}$$

where $\omega(x)$ is any positive scalar function, viz.

$$\frac{d}{dt} l(x(t)) \leqslant 0, \quad \frac{d}{dt} d(x^* \| x(t)) \leqslant 0$$

for suitable x^*.

4. MONOTONICITY OF THE EXPLICIT ALGORITHM

In this section we prove the monotomicity results for the explicit algorithm. They are slightly messier than for the implicit case.

Let $G: \mathbb{R}_+^N \to \mathbb{R}_+^N$ be defined as

$$[G(x)]_j = x_j / (1 + \omega(x)[\nabla l(x)]_j), \; j = 1, 2, \ldots, N. \tag{4.1}$$

By the choice of $\omega(x)$ obviously $G(x) > 0$ if $x > 0$, and $[G(x)]_j = 0$ only if $x_j = 0$.

LEMMA 4.1. *For all $x \geqslant 0$ we have $l(x) \geqslant l(G(x))$ with equality if and only if $x = G(x)$.*

PROOF. From the convexity of l we have

$$l(x) - l(G(x)) \geqslant < \nabla l(G(x)), x - G(x) >,$$

which equals

$$<\nabla l(x),\ x-G(x)> - <\nabla l(x)-\ \nabla l(G(x)),\ x-G(x)>$$

Since ∇l is locally Lipschitz continuous, this expression is greater than or equal to

$$<\nabla l(x),\ x-G(x)> - L\|x-\ G(x)\|_2^2 =$$

$$\omega(x)\sum_j \frac{x_j|[\nabla l(x)]_j|^2}{1+\omega(x)[\nabla l(x)]_j}\left[1-\frac{\omega(x)\,Lx_j}{1+\omega(x)[\nabla l(x)]_j}\right],$$

for the appropriate constant L. By the choice of $\omega(x)$ this dominates

$$\tfrac{1}{4}\,\omega(x)\sum_j x_j\,|[\nabla l(x)]_j|^2 = \tfrac{1}{4}\,\omega(x) <\nabla l(x), D(x)\,\nabla l(x)>.$$

and the conclusion follows as in the proof of lemma 3.1. □

We state some consequences of this lemma, before proceeding with the second monotonicity result.

COROLLARY 4.2. *$\{l(x^n)\}_n$ is decreasing.*

COROLLARY 4.3. *$\{x^n\}_n$ is bounded, and every subsequence itself has a convergent subsequence.*

PROOF. Since $l(x^n) \geqslant l(x^{n+1})$ then for every n,

$$x^n \in \{x \geqslant 0 : l(x) \leqslant l(x^0)\}$$

which is a compact set. □

COROLLARY 4.4. *Let $V(x) = \omega(x) <\nabla l(x), D(x)\,\nabla l(x)>$. Then*

$$V(x) \leqslant 4[l(x) - l(G(x))].$$

This last result just summarizes the proof of lemma 4.1.

We now introduce the following notation. For $x \in \mathbb{R}_+^N$ and $y > 0$ let

$$e(x\|y) = d(x\|y) + 8\,[l(y) - l(x)]/M \tag{4.4}$$

Note that $e(x\|y)$ is a measure for $\|x-y\|_2$, *provided* x solves the minimization problem (1.1), or is otherwise known to satisfy $l(x) \leqslant l(y)$.

Let m and Ω be defined by (3.2)-(3.3) with $\{x^n\}_n$ generated by the explicit algorithm.

LEMMA 4.4. *Let $x^* \in \Omega$. Then $e(x^*\|x^n) \geqslant e(x^*\|x^{n+1})$ for all n.*

PROOF. Since $x^0 > 0$ we have $x^n > 0$ for all n, and so $d(x^*\|x^n)$ is well defined. Again

$$d(x^*\|x^n) - d(x^*\|x^{n+1}) = \sum_j x_j^* \log(x_j^{n+1}/x_j^n) + x_j^n - x_j^{n+1} =$$

$$= \sum_j x_j^* \log(1+\omega_n[\nabla(x^n)]_j)^{-1} + \omega_n x_j^n[\nabla l(x^n)]_j/(1+\omega_n[l(x^n)]_j).$$

Once more, $\log(1+t)^{-1} \geqslant -t$ and so the above expression dominates

$$\omega_n <\nabla l(x_n),\ x^n - x^*> - \sum_j x_j^n|\omega_n[\nabla l(x^n)]_j|^2/(1+\omega_n[\nabla l(x^n)]_j).$$

By the choice of ω_n the last sum is dominated by $2\omega_n V(x^n)$ (see Corollary 4.4) and thus by $8[l(x_n) - l(x^{n+1})]/M$. We have thus shown that

$$d(x^*\|x^n) - d(x^*\|x^{n+1}) \geqslant \omega_n[l(x^n) - l(x^*)] - 8[l(x^n) - l(x^{n+1})]/M.$$

Since $l(x^n) \geqslant l(x^*)$ the result follows. □

5. Convergence of the explicit algorithms

With the Lemmas 3.1-3.2 and 4.1-4.4 in hand, the convergence proofs of the two algorithms are virtually identical. We will just consider the explicit algorithms, since this is the slightly more complicated case.

LEMMA 5.1. *The sequence $\{x^n\}_n$ generated by the explicit algorithm converges.*

PROOF. By Corollary 4.3 the sequence $\{x^n\}_n$ has a convergent subsequence $\{x^{n_k}\}_k$ with limit x^*, say. Then $l(x^*) = \lim_{k\to\infty} l(x^{n_k})$. We may now apply Lemma 4.4 to our x^*, thus $e(x^* \| x^n) \geqslant e(x^* \| x^{n+1})$ for all n. Now for our subsequence $\{x^{n_k}\}$ we have $e(x^* \| x^{n_k}) \to 0$ as $k \to \infty$, and by the monotonicity then

$$\lim_{n\to\infty} e(x^* \| x^n) = 0.$$

Then also $d(x^* \| x^n) \to 0$, and the convergence of $\{x^n\}_n$ to x^* follows. □

We finally need to show that the x^* obtained above solves the minimization problem (1.1). First we have that x^* is a fixed point of the iteration $x^{n+1} = G(x^n)$, so that

$$x_j^* \, [\nabla l(x^*)]_j = 0\,,\ j = 1,2,\ldots,N. \tag{5.1}$$

We have of course also that

$$x_j^* \geqslant 0\,,\ j = 1,2,\ldots,N. \tag{5.2}$$

In order to show that $\nabla l(x^*) \geqslant 0$ consider for a fixed j the ratios $r^n = x_j^{n+1}/x_j^n$. Since $\{x^n\}_n$ is bounded it follows that $\liminf_{n\to\infty} r^n \leqslant 1$ or

$$\liminf_{n\to\infty} \omega_n [\nabla l(x^n)]_j \geqslant 0.$$

It follows that $[\nabla l(x^*)]_j \geqslant 0$. Since j was arbitrary,

$$[\nabla l(x^*)]_j \geqslant 0\,,\ j = 1,2,\ldots,N. \tag{5.3}$$

Thus, x^* satisfies the Kuhn-Tucker conditions, and is a solution of the minimization problem (1.1). We have thus proven

THEOREM 5.2. *The sequence $\{x^n\}_n$ generated by the explicit algorithm starting from an $x^0 > 0$ converges to a solution of the minimization problem (1.1).*

It should be remarked that in the above proof we can show that x^* satisfies the Kuhn-Tucker conditions only after we have shown that $\{x^n\}_n$ converges to x^*. In the absence of Lemma 4.4 we do not have this convergence, and indeed it does not appear to be easy to prove that x^* is optimal only from knowing that a subsequence of $\{x^n\}$ converges to x^* (so $l(x^*) = \lim_{n\to\infty} l(x^n)$), without some further assumptions on $l(x)$.

6. Applications to linear models

In this section we discuss some applications of multiplicative iterative algorithms. We consider the following class of convex programming problems. Suppose some physical phenomenon is modelled by the system of linear equations

$$Ax = b \tag{6.1}$$

where $A \in \mathbb{R}^{m\times N}$ is a nonnegative matrix with nonzero column sums, the data vector $b \in \mathbb{R}^M$ is nonnegative and $x \in \mathbb{R}^N_+$ describes the physical quantity to be identified. Due to all kinds of approximations in the modelling process (linearization, statistical effects) the notion of a good solution of the system (6.1) needs to be clarified. A common choice is to interpret (6.1) in the least squares sense. The "solution" of (6.1) is defined to be the solution of the minimization problem

$$\begin{aligned} &\text{minimize } \|Ax-b\|_2^2 \\ &\text{subject to } x \geqslant 0. \end{aligned} \tag{6.2}$$

Slightly more general, let $L(y)$ be a strictly convex function on $\mathbb{R}^M_+$, with compact level sets and locally Lipschitz continuous gradient. Typically, L will also depend on the data vector b, but we suppress this dependence in the notation. Now choose as interpretation of (6.1)

$$\begin{aligned} &\text{minimize } L(Ax) \\ &\text{subject to } x \geqslant 0\,, \end{aligned} \tag{6.3}$$

and we are interested in applying the explicit multiplicative iterative algorithm to the function $l(x)=L(Ax)$. We first show that l satisfies the assumptions if Section 2. First of all, it is easy to show that $l(x)$ is convex, and has a locally Lipschitz continuous gradient. It should be remarked that $l(x)$ is not strictly convex if A has a nontrivial nullspace. There remains the question of compact level sets of $l(x)$. To settle this, note that for fixed $y \in \mathbb{R}^N_+$ the level set of L

$$\{z \in \mathbb{R}^M_+ \, : \, L(z) \leqslant L(y)\}$$

is compact, and so is its intersection with

$$A(\mathbb{R}^N_+) = \{Ax \, : \, x \in \mathbb{R}^N_+\}.$$

Thus it follows that for every $z \in \mathbb{R}^N_+$ the set

$$\{Ax \, : \, l(x) \leqslant l(z)\,,\; x \in \mathbb{R}^N_+\} \tag{6.4}$$

is compact. Since A is nonnegative and has nonzero column sums we have that $\mathfrak{N}(A) \cap \mathbb{R}^N_+ = \{0\}$, where $\mathfrak{N}(A) = \{x \in \mathbb{R}^N : Ax = 0\}$ is the nullspace of A, and so the compactness of (6.4) implies the compactness of

$$\{x \in \mathbb{R}^N_+ \, : \, l(x) \leqslant l(z)\},$$

i.e. l has compact level sets. Thus l satisfies all our assumptions. The conclusion is that the algorithm

$$\begin{aligned} &x^0 > 0 \\ &x_j^{n+1} = x_j^n / (1 + \omega_n [A^T \nabla L(Ax^n)]_j), \; j = 1,2,\ldots,N\,, \end{aligned}$$

converges to a solution of the minimization problem (6.3).

We now discuss a few specific examples.

6.1 Constrained least squares

We already mentioned the minimization problem (6.1). It is not clear whether the original multiplicative algorithm (1.3) converges in case $\mathfrak{N}(A) \neq \{0\}$. In case $\mathfrak{N}(A) = \{0\}$ the solution of (6.2) is unique, and it is not hard to show that then convergence follows. The explicit multiplicative algorithm

$$x_j^{n+1} = x_j^n / (1 + \omega_n [A^T(Ax^n - b)]_j\,, \; j = 1,2,\ldots,N\,, \tag{6.5}$$

converges for appropriate ω_n, whether A has full column rank or not.

6.2 Image space reconstruction algorithm (ISRA)

If we apply a diagonal scaling in the constrained least squares problem we get

$$\begin{aligned} &\text{minimize } \|AD(\lambda)y - b\|_2^2 \\ &\text{subject to } y \geqslant 0 \end{aligned} \tag{6.6}$$

where $D(\lambda)$ is a diagonal matrix with diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_N$ which are all strictly positive, and we take $x = D(\lambda)y$ as our solution of (6.1). The algorithm (6.5) applied to (6.6) gives the iteration formula

$$y_j^{n+1} = y_j^n / (1 + \omega_n \lambda_j [A^T(AD(\lambda)y^n - b)]_j), \; j = 1,2, \ldots, N,$$

and scaling back to x^n,

$$x_j^{n+1} = x_j^n / (1 + \omega_n \lambda_j [A^T(Ax^n - b)]_j), \; j = 1,2, \ldots, N \tag{6.7}$$

In emission tomography where the column sums of A are normalized to one,

$$\sum_{i=1}^{M} a_{ij} = 1, \; j = 1,2, \ldots, N,$$

the above algorithm is applied with $\omega_n = 1$ and $\lambda_j = [A^T b]_j$. The resulting algorithm has the elegant form

$$x_j^{n+1} = x_j^n [A^T b]_j / [A^T A x_n]_j, \; j = 1,2, \ldots, N \tag{6.8}$$

This algorithm is the image space reconstruction algorithm (ISRA) of Daube-Witherspoon and Muehllehner [2]. De Pierro [3] proves that for algorithm (6.8) we have

$$\|Ax^n - b\|_2^2 \geq \|Ax^{n+1} - b\|_2^2, \tag{6.9}$$

and proves convergence when the solution of (6.2) is unique. When this is not so, the convergence of this method follows from the general theory of Section 5 and Lemma 4.4, combined with (6.9).

It should be remarked that algorithm (6.8) has been applied in remote sensing for the reconstruction of temperature profiles, see Chu[1] and references therein, but then in the form, cf. equation (6.1),

$$x_j^{n=1} = x_j^n \, b_j / [Ax^n]_j \, , \; j = 1,2, \ldots, N. \tag{6.10}$$

This must be regarded with even more scepticism than the algorithm (6.8), since the above algorithm necessarily generates an oscillating sequence $\{x^n\}_n$ if the system (6.1) is not consistent. The same criticism would apply to the variation proposed by Twomey [9].

Note that De Pierro [3] considers the minimization problem

$$\text{minimize } \frac{1}{2} \langle x, Mx \rangle - \langle b, x \rangle$$

$$\text{subject to } x \geq 0$$

with M positive definite and $M \geq 0$ element wise. Our theory applies to this case as well, even when M is only nonnegative definite (but still $M \geq 0$ element wise) since then the level sets remain compact.

6.3 Maximum likelihood estimation in emission tomography

In emission tomography the problem can be stated as the minimization of $l(x) = d(b \| Ax)$ over $\mathbb{R}_+^N$. Here $d(\cdot \| \cdot)$ is the Kullback-Leibler divergence (1.6). The *EM* algorithm of Shepp and Vardi [8] has the simple form

$$x_j^{n+1} = x_j^n [A^T r^n]_j \, , \; j = 1,2, \ldots, N, \tag{6.11}$$

with $r_i^n = b_i / [Ax^n]_i$, $i = 1,2, \ldots, M$. It is a special case of algorithm (1.2), in this case

$$x_j^{n+1} = x_j^n \{1 - \omega_n (1 - [A^T r^n]_j)\} \, , \; j = 1,2, \ldots, N. \tag{6.12}$$

Vardi et al. [10] were able to prove Lemma 4.1 (easy) and Lemma 4.2 (hard) in this case, but it only works for $0 < \omega_n \leq 1$. However, Lewitt and Muehllehner [6] in their experiments advocated the choice $\omega_n \approx 4$ after a few iterations so convergence in this case is still an open question. In our version of the algorithm we would have

$$x_j^{n+1} = x_j^n / \{1 + \omega_n (1 - [A^T r^n]_j)\}, \; j = 1,2, \ldots, N, \tag{6.13}$$

and there are no restraints on the size of ω_n other than the (easily modifiable) choice (2.4), see the comments on this choice in section 2. It remains of course to be seen what the practical difference (if any) is between the algorithms (6.12) and (6.13).

REFERENCES

[1] W.P. CHU. Convergence of Chahine's nonlinear relaxation inversion method used for limb viewing remote sensing. *Applied Optics 24:* 445-447 (1985).

[2] M.E. DAUBE-WITHERSPOON and G. MUEHLLEHNER. An iterative image space reconstruction algorithm suitable for volume ECT. *IEEE Trans. Medical Imaging MI-5:* 61-66 (1986).

[3] A.R. DE PIERRO. On the convergence of the iterative image space reconstruction algorithm for volume ECT. *IEEE Trans. Medical Imaging. MI-6:* 174-175 (1987).

[4] F. LIESE and I. VAJDA. Convex statistical distances. Leipzig: Teubner (1987).

[5] O.L. MANGASARIAN. Nonlinear programming. New York: McGraw-Hill (1969).

[6] R.M. LEWITT and G. MUEHLLEHNER. Accelerated iterative reconstruction for position emission tomography based on the EM algorithm for maximum likelihood estimation *IEEE Trans. Medical Imaging MI-5:* 16-22 (1986).

[7] R.T. ROCKAFELLAR. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimiz. 14:* 877-898 (1976).

[8] L.A. SHEPP and Y. VARDI. Maximum likelihood reconstruction in emission tomography. *IEEE Trans. Medical Imaging MI-1:* 113-122 (1982).

[9] S. TWOMEY. Comparison of constrained linear inversion and an iterative nonlinear algorithm applied to the indirect estimation of particle size distribution. *J. Comput Phys, 18:* 188-200 (1975).

[10] Y. VARDI, L.A. SHEP and L. KAUFMAN. A statistical model for position emission tomography. *J. Amer. Statist. Assoc. 80:* 8-38 (1985).