ITERATIVE RECONSTRUCTION ALGORITHMS BASED ON CROSS-ENTROPY MINIMIZATION

CHARLES BYRNE*

Abstract. The "expectation maximization maximum likelihood" algorithm (EMML) has received considerable attention in the literature since its introduction in 1982 by Shepp and Vardi. A less well known algorithm, discovered independently in 1972 by Schmidlin ("iterative separation of sections") and by Darroch and Ratcliff ("generalized iterative scaling"), and rediscovered and called the "simultaneous multiplicative algebraic reconstruction technique" (SMART) in 1992, is quite similar to the EMML. Both algorithms can be derived within a framework of alternating minimization of cross-entropy distances between convex sets. By considering such a parallel development of EMML and SMART we discover that certain questions answered for SMART remain open for EMML. We also demonstrate the importance of cross-entropy (or Kullback-Leibler) distances in understanding these algorithms, as well as the usefulness of Pythagorean-like orthogonality conditions in the proofs of the results. The SMART is closely related to the "multiplicative algebraic reconstruction technique" (MART) of Gordon, Bender and Herman; we include a derivation of MART within the same alternating minimization framework and provide an elementary proof of the convergence of MART in the consistent case, extending the theorem of Lent. Some partial results on the behavior of MART in the inconsistent case are also discussed.

1. Introduction. In 1976 Rockmore and Macovski [1] suggested that the statistical model of independent Poisson emitters play a more explicit role in emission tomography image reconstruction and that the vector of spatially distributed Poisson means be estimated using the well known "maximum likelihood" (ML) procedure. In their 1982 paper [2] Shepp and Vardi adopted this suggestion and employed the iterative "expectation maximization " (EM) algorithm, presented in a more general setting by Dempster, Laird and Rubin [3], to obtain the ML solution. Working independently, Lange and Carson [4] obtained some convergence results for the EMML. A more complete treatment of convergence was given in [5], using results of [6]. Regularization of the EMML using penalty functions derived from a Bayesian framework was included in [7]. In [8] Titterington noted that the EMML leads to an iterative algorithm for finding nonnegative solutions of the linear system $y = Px$, where $y \geq 0$, $P \geq 0$ is an I by J matrix with column sums 1 and $x \geq 0$ is sought. In the inconsistent case, in which there are no nonnegative solutions of $y = Px$, the algorithm produces a nonnegative minimizer of $KL(y, Px)$; here $KL(a, b)$ denotes the (nonsymmetric) Kullback-Leibler or cross-entropy distance between the nonnegative vectors a and $b : KL(a, b) = \sum a_n \log(a_n/b_n) + b_n - a_n \geq 0$.

A second algorithm, discovered independently by Schmidlin [9,10] and by Darroch and Ratcliff [11,12] in 1972, and rediscovered in 1992 and named the "simultaneous MART" (SMART) algorithm [13,14], leads to a nonneg-

* Department of Mathematics, University of Massachusetts Lowell, Lowell, MA 01854.

to APPEAR IN PROC. of IMA WORKSHOP ON IMAGE PROCESSING, MAY 94
(L. SHEPP, EDITOR) SPRINGER.

ative minimizer of $KL(Px, y)$. The EMML and SMART algorithms are similar in many respects, but the nonsymmetric nature of the KL distance leads to curious divergences in the theoretical development, with the result that certain questions that have been answered for SMART remain unsolved for EMML.

The SMART algorithm is related to the "multiplicative algebraic reconstruction technique" (MART) [15] as the Jacobi method is related to Gauss-Seidel; MART updates the estimate after each calculation involving a single equation, whereas SMART updates the estimate only after all the equations have been considered. The derivations of SMART and MART within a framework of cross-entropy minimization are quite similar and the algorithms give the same solution in the consistent case. In all the simulations of the inconsistent case we have considered, the MART behaves the same way the ART does, converging, not to a single vector, but to a limit cycle of as many vectors as there are equations. It remains to prove that this always happens and to uncover properties of this limit cycle.

The purpose of this article is to present the development of these algorithms in a way that highlights both the similarities and the differences, as well as to illustrate the important role of the KL distance in this development. In what follows results pertaining to one of the algorithms will be labeled accordingly. We first develop the EMML and SMART in parallel, and then treat the MART.

2. The EMML and SMART algorithms. EMML In emission tomography the EMML is typically applied in the inconsistent case, in which, because of additive noise, there is no nonnegative solution of $y = Px$. So the algorithm is viewed primarily as a minimizer of $KL(y, Px)$. Either $I \geq J$ or $I \leq J$, and both are usually large.

SMART Darroch and Ratcliff [11] designed their algorithm to select one probability vector out of the many satisfying a given set of linear constraints; typically $I < J$. For example, we might want a probability vector on a product space having prescribed marginals. So the SMART was initially viewed as minimizing $KL(x, x^0)$ over all $x \geq 0$ with hard constraints $y = Px$. In the inconsistent case, it turns out that SMART minimizes $KL(Px, y)$ [13,14], hence the connection with EMML.

In what follows we place no limitations on I and J and consider the behavior of both algorithms in the consistent and the inconsistent cases. By considering the algorithms in parallel we can see how arguments valid for SMART break down when applied to EMML.

2.1. Projecting onto convex sets. Let $R = \{r = r(x) = \{r_{i,j} = x_j P_{i,j} y_i / Px_i\}\}$, so that $\sum r_{i,j} = y_i$ for each fixed i.Let $Q = \{q = q(x) = \{q_{i,j} = P_{i,j} x_j\}\}$. Notice that if we can find a $q(x)$ in the intersection of R and Q then $y = Px$, so we have found a nonnegative solution. In both algorithms below the idea is to alternately project onto each of the two sets; only the distances involved differ. For both EMML and SMART we begin with $x^0 > 0$.

EMML 1) minimize $KL(r(x), q(x^k))$ to get $x = x^k$;
2) minimize $KL(r(x^k), q(x))$ to get $x = x^{k+1}$;

then **the EMML ALgorithm is** $x_j^{k+1} = x_j^k \sum P_{i,j}(y_i/Px_i^k)$, for each $j = 1, ..., J$ and for $k = 0, 1, 2, ...$

SMART 1) minimize $KL(q(x^m), r(x))$ to get $x = x^m$;
2) minimize $KL(q(x), r(x^m))$ to get $x = x^{m+1}$;

then **the SMART algorithm is** $x_j^{m+1} = x_j^m \exp[\sum P_{i,j} \log(y_i/Px_i^m)]$, for each $j = 1, ..., J$ and for $m = 0, 1, 2, ...$

The convex sets R and Q occur in the convergence proofs in [5] and in [6,12].

2.2. Orthogonality. In each of the minimizations above a sort of orthogonal projection is taking place. Associated with each projection is a "Pythagorean identity" involving the KL distance, expressing the underlying orthogonality. It is best to think of KL as distance squared for the purpose of understanding these identities.

EMML 1 $KL(r(x), q(x^k)) = KL(r(x^k), q(x^k)) + KL(r(x), r(x^k))$.

EMML 2 $KL(r(x^k), q(x)) = KL(r(x^k), q(x^{k+1})) + KL(x^{k+1}, x)$.

SMART 1 $KL(q(x^m), r(x)) = KL(q(x^m), r(x^m)) + KL(x^m, x) - KL(Px^m, Px)$.

SMART 2 $KL(q(x), r(x^m)) = KL(q(x^{m+1}), r(x^m)) + KL(x, x^{m+1})$.

Note: we have that, for all x and z,

$$KL(x, z) - KL(Px, Pz) = \Sigma\Sigma KL(P_{i,j}x_j, P_{i,j}z_j Px_i/Pz_i) \geq 0. \tag{2.1}$$

Note: since $KL(r(x), q(x)) = KL(y, Px)$ and $KL(q(x), r(x)) = KL(Px, y)$, the functions to be minimized with respect to x to get x^{k+1} and x^{m+1} are

$$KL(r(x^k), q(x)) = KL(y, Px) + KL(r(x^k), r(x)), \tag{2.2}$$

for EMML, and, for SMART,

$$KL(q(x), r(x^m)) = KL(px, y) + KL(x, x^m) - KL(Px, Px^m). \tag{2.3}$$

Since the functions we wish ultimately to minimize are $KL(y, Px)$ and $KL(Px, y)$ respectively, we can view (2.2) and (2.3) as sequential unconstrained methods involving barrier or penalty functions in the sense of [16].

The problem of minimizing a function $f(x)$, subject to the nonnegativity constraints $x \geq 0$, will play a central role in what follows. The

Kuhn-Tucker conditions [16] are necessary for $x \geq 0$ to be a global minimizer of $f(x)$:

$$\partial f(x)/\partial x_j = 0, \qquad \text{if } x_j > 0; \tag{2.4}$$

$$\partial f(x)/\partial x_j \geq 0, \qquad \text{if } x_j = 0. \tag{2.5}$$

Here $\partial f(x)/\partial x_j$ denotes the first partial derivative of f, with respect to the jth entry of x, evaluated at the vector x. The functions f we shall be considering are convex; for such functions (2.4) and (2.5) are also sufficient for x to be a global minimizer.

2.3. Necessary conditions for minimizers. Applying (2.4) to the function $f(x) = KL(y, Px)$ for EMML, and to the function $f(x) = KL(Px, y)$ for SMART, we obtain (2.6) and (2.7) respectively, for any $x \geq 0$ minimizing $f(x)$:

$$1 = \Sigma P_{i,j}(y_i/Px_i), \text{ for all } j \text{ such that } x_j > 0; \tag{2.6}$$

$$1 = \exp[\Sigma P_{i,j} \log(y_i/Px_i)] \text{for all } j \text{ such that } x_j > 0. \tag{2.7}$$

The EMML and SMART iterations can then be derived from (2.6) and (2.7); multiply both sides by x_j to obtain a fixed-point equation valid for all x_j and then use the most recent estimate of x on the right side to produce the next estimate of x on the left. Of course a more sophisticated derivation is required if convergence is to be established.

EMML 3 The sequence $\{x^k\}$ is contained within a compact set.

Proof. $\Sigma x_j^k = \Sigma y_i$ for each k. □

SMART 3 The sequence $\{x^m\}$ is contained within a compact set.

Proof. $\Sigma x_j^m \leq \Sigma y_i$ for each k. □

We shall denote by x^* an arbitrary subsequential limit of either $\{x^k\}$ or $\{x^m\}$.

EMML 4 $\{KL(y, Px^k)\}$ is decreasing and so $\{KL(x^{k+1}, x^k)\}$ is going to zero.

Proof. We have $KL(y, Px^k) = KL(y, Px^{k+1}) + KL(r(x^k), r(x^{k+1})) + KL(x^{k+1}, x^k)$ from EMML 1 and EMML 2. □

SMART 4 $\{KL(Px^m, y)\}$ is decreasing and so $\{KL(x^m, x^{m+1})\}$ is going to zero.

Proof. The argument is similar to that of EMML 4. □

EMML-SMART 5 For either of the two iterative schemes above, let x' denote the next term in the iteration following x. Then $(x^*)' = x^*$, that is any subsequential limit point is a fixed point of the iteration.

We turn not the the consistent case, in which nonnegative solutions of $y = Px$ exist.

2.4. Consistent case. Throughout this subsection we assume that $x \geq 0$ satisfies $y = Px$.

EMML 6 $KL(x, x^k) - KL(x, x^{k+1}) \geq KL(y, Px^k) \geq 0$, so $\{KL(y, Px^k)\}$ goes to zero.

Proof. The left side is $\Sigma x_j \log(\Sigma P_{i,j} y_i / P x_i) \geq \Sigma x_j (\Sigma P_{i,j} \log(y_i / P x_i)) = KL(y, Px^k)$, since $\Sigma x_j^k = \Sigma x_j^{k+1}$. □

<u>SMART 6</u> $KL(x, x^m) - KL(x, x^{m+1}) \geq KL(Px^m, y) \geq 0$, so $\{KL(Px^m, y)\}$ goes to zero.

Proof. $KL(x, x^m) = KL(x, x^{m+1}) + KL(Px^m, y) + KL(q(x^{m+1}), r(x^m))$. □

<u>EMML-SMART 7</u> $y = Px^*$

<u>EMML 8</u> [2,4,5,6,7,13] The sequence $\{x^k\}$ converges to $x^* = x^\infty$, $KL(x, x^\infty) < \infty$, so the support of x^∞ is maximal over all solutions x.

Proof. Use $x = x^*$ in EMML 6. Then $\{KL(x^*, x^k)\}$ is decreasing. □

<u>SMART 8</u> [9,10,11,12,13,14] The sequence $\{x^m\}$ converges to $x^* = x^\infty$; $KL(x, x^\infty) < \infty$, so the support of x^∞ is maximal over all solutions x. In addition, x^∞ is the unique solution for which $KL(x, x^0)$ is minimized. If x^0 is a constant vector then x^∞ is the maximum Shannon entropy solution.

Proof. The first part is similar to the proof of EMML 8. For any solution x we have $KL(x, x^m) - KL(x, x^{m+1}) = KL(Px^m, y) + KL(q(x^{m+1}), r(x^m))$, which is independent of the x chosen. Consequently, by "telescoping", $KL(x, x^0) - KL(x, x^\infty)$ is also independent of x. Therefore, minimizing $KL(x, x^0)$ over all solutions $x \geq 0$ is equivalent to minimizing $KL(x, x^\infty)$ over the same x; but the solution to the latter problem is obviously x^∞. □

<u>EMML</u> The difference $KL(x, x^k) - KL(x, x^{k+1})$ is not independent of x, so the proof above breaks down when applied to the EMML case. It remains an open question to which solution the EMML converges in this case.

2.5. Inconsistent case. We assume now that there are no $x \geq 0$ for which $y = Px$.

LEMMA 2.1. *For any nonnegative vectors a and b we have $KL(a, b) \geq KL(\Sigma a_n, \Sigma b_n)$.*

Proof. Minimize the left side subject to equality constraints on Σa_n and Σb_n. □

<u>EMML 9</u> Let $x \geq 0$ be any nonnegative minimizer of $KL(y, Pz)$. Then $x' = x$ and we have $KL(x, x^k) \geq KL(r(x), r(x^k)) \geq KL(x, x^{k+1})$, so that $\{KL(x, x^k)\}$ is decreasing and $KL(x, x^*) < \infty$. Therefore $KL(x, x^*) \geq KL(r(x), r(x^*)) \geq KL(x, x^*)$, so that $KL(x, x^*) = KL(r(x), r(x^*))$.

Proof. That $x' = x$ follows from the Kuhn-Tucker theorem and the definition of the iteration. The second inequality follows from the lemma. The first inequality is a little harder. We have

$$KL(r(x), q(x^k)) = KL(r(x), q(x)) + KL(x, x^k) = KL(y, Px) + KL(x, x^k),$$

and also

$$\begin{aligned} KL(r(x), q(x^k)) &= KL(r(x^k), q(x^k)) + KL(r(x), r(x^k)) = KL(y, Px^k) + \\ &\quad + KL(r(x), r(x^k)). \end{aligned}$$

Since $KL(y, Px) \leq KL(y, Px^k)$ the result follows. □

SMART 9 Let $x \geq 0$ be any nonnegative minimizer of $KL(Pz, y)$. Then $x' = x$ and we have $KL(x, x^m) \geq KL(x, x^{m+1})$, so that $\{KL(x, x^m)\}$ is decreasing and $KL(x, x^\infty) < \infty$.

Proof. That $x' = x$ follows from the Kuhn -Tucker theorem and the definition of the iteration. We have

$$\begin{aligned} KL(x, x^m) - KL(x, x^{m+1}) = \\ KL(Px^{m+1}, y) - KL(Px, y) + KL(Px, Px^m) + KL(x^{m+1}, x^m) - \\ -KL(Px^{m+1}, Px^m) \geq 0; \end{aligned}$$

note that $KL(Px^{m+1}, y) - KL(Px, y) \geq 0$ by the choice of x. □

EMML 10 [2,4,5,6,7,13] $Px^* = Px$, so x^* is a global nonnegative minimizer of $KL(y, Pz)$. So $\{KL(x^*, x^k)\}$ is decreasing and $\{x^k\}$ converges to $x^* = x^\infty$.

Proof. We have

$$KL(r(x), q(x^*)) = KL(y, Px) + KL(x, x^*)$$

and

$$KL(r(x), q(x^*)) = KL(y, Px^*) + KL(r(x), r(x^*)) = KL(y, Px^*) + KL(x, x^*).$$

Since $KL(x, x^*) < \infty$, $KL(y, Px^*) = KL(y, Px)$ follows. By the strict convexity of KL the optimal Px is unique, even if x is not, so $Px^* = Px$. The other assertion follows by using x^* in place of x in EMML 9. □

SMART 10 [13,14] $Px^* = Px$, so that x^* is a global nonnegative minimizer of $KL(Pz, y)$. Therefore $\{KL(x^*, x^m)\}$ is decreasing and so $\{x^m\}$ converges to $x^* = x^\infty$. In addition, x^∞ is the unique nonnegative minimizer of $KL(Px, y)$ for which $KL(x, x^0)$ is minimized.

Proof. From the proof of SMART 9 we have that $KL(x, x^m) - KL(x, x^{m+1}) \geq KL(Px, Px^m)$, so it follows that $\{KL(Px, Px^m)\}$ converges to zero; hence $KL(Px, Px^*) = 0$. Now use x^* as x in SMART 9. To prove the last assertion, we not that $KL(x, x^m) - KL(x, x^{m+1})$ is again independent of x (but not of Px, which is unique, even if x is not). Use the "telescoping" argument as in the proof of SMART 8. □

EMML Note that the distance $KL(x, x^k) - KL(x, x^{k+1})$ is not independent of x, so that same argument fails for EMML. It remains an open question to which nonnegative minimizer of $KL(y, Px)$ the EMML converges.

2.6. Uniqueness of the solution in the inconsistent case. Again let x denote a nonnegative minimizer of $KL(y, Px)$ or of $KL(Px, y)$ as appropriate. If we assume that P has the "full rank property" then x is unique in both cases.

Definition Say that the matrix P has the "full rank property" (FRP) if P and all the matrices Q obtained from P by deletion of columns have full rank.

If we assume the FRP for P, we obtain the following uniqueness and support results:

EMML 11 [13] There is a subset S of $\{1, 2, ..., J\}$, having cardinality at most $I - 1$, such that, for each nonnegative minimizer x of $KL(y, Px)$, we have $x_j > 0$ only if j is in S. Therefore there is but one such x, and it has fewer than I nonzero entires.

Proof. From $x' = x$ it follows that if $x_j > 0$, then $\Sigma P_{i,j}(y_i/Px_i) = 1$. Let $u = (u_i) = (y_i/Px_i)$; note that Px is unique, even if x is not. Let Q be obtained from P by deleting the jth column whenever $x_j = 0$ for all x. Then we have $Q^T u = 1 = (1, ..., 1)^T$. But, from the assumption that the columns of P sum to 1 it follows that $Q^T 1 = 1$. Since Q has full rank, it follows that if Q has I or more columns then Q is one-to-one, hence $u = 1$, or the system $y = Px$ is satisfied. □

SMART 11 There is a subset S of $\{1, 2, ..., J\}$, having cardinality at most $I - 1$, such that for each nonnegative minimizer x of $KL(Px, y)$, we have $x_j > 0$ only if j is in S. Therefore there is but one such x, and it has fewer than I nonzero entries.

Proof. The proof is essentially the same as that of EMML 11. □

Comment We have not established that the sets S in the two theorems are the same, although this has been the case in all simulations we have performed.

3. The MART algorithm. We can obtain the MART algorithm [15] by using an alternating minimization approach similar to that used for SMART above. However, each step of the MART iteration involves only a single equation; therefore we must minimize a different distance each time, depending on which equation is being used at that step. The distance associated with the ith equation is

$$G_i(z, x) = KL(z, x) - KL(Pz_i, Px_i) + KL(Pz_i, y_i). \tag{3.1}$$

We know that $KL(z, x) - KL(Pz, Px) \geq 0$, so $G_i(z, x) \geq 0$. Beginning with $z^0 > 0$, the MART iterative scheme is obtained as follows: for $m = 0, 1, 2, ...$ and $i = m(mod I) + 1$, perform

Step 1): minimize $G_i(z^m, x)$ to get $x = z^m$;

Step 2): minimize $G_i(z, z^m)$ to get $z = z^{m+1}$;

then the MART algorithm is

$$z_j^{m+1} = z_j^m \exp[P_{i,j} \log(y_i/Pz_i^m)] = z_j^m (y_i/Pz_i^m)^{P_{i,j}} \tag{3.2}$$

for each $j = 1, ..., J$ and for $m = 0, 1, 2, ...$

3.1. MART in the consistent case. We prove the convergence of the MART algorithm in the consistent case in a manner similar to that used for SMART. Directly from the definition we have

MART 1 $G_i(z^m, x) = KL(z^m, x) - KL(Pz_i^m, Px_i) + KL(Pz_i^m, y_i)$.

MART 2 $G_i(x, z^m) = G_i(z^{m+1}, z^m) + KL(x, z^{m+1})$.

Proof. Calculate $G_i(x, z^m) - G_i(z^{m+1}, z^m)$. □

Assume that $x \geq 0$ is such that $y = Px$. We then have

MART 3 The sequence $\{KL(x, z^m)\}$ is decreasing and so $\{G_i(z^{m+1}, z^m)\}$ and $\{KL(y_i, Pz_i^m)\}$ converge to zero and $\{z^m\}$ is contained within a bounded set.

Proof. From the definition of $G_i(z, x)$ and MART 1,2 we have

$$KL(x, z^m) = G_i(x, z^m) + KL(Px_i, Pz_i^m) - KL(Px_i, y_i) = G_i(x, z^m) + +KL(Px_i, Pz_i^m)$$

so

$$KL(x, z^m) - KL(x, z^{m+1}) = G_i(z^{m+1}, z^m) + KL(y_i, Pz_i^m).$$

□

We prove the following theorem:

MART 4 When there are nonnegative solutions of $y = Px$ the MART sequence $\{z^m\}$ converges to the unique solution for which $KL(x, z^0)$ is minimized. If z^0 is a constant vector, then $\{z^m\}$ converges to the solution maximizing the Shannon entropy.

Remark Lent [17] has shown convergence to the maximum Shannon entropy solution for the MART case and for z^0 a constant vector.

Proof. From MART 1 $G_i(z^{m+1}, z^m) = KL(z^{m+1}, z^m) - KL(Pz_i^{m+1}, Pz_i^m) + KL(Pz_i^{m+1}, y_i)$. Since $\{KL(y_i, Pz_i^m)\}$ goes to zero, it follows that $\{KL(Pz_i^{m+1}, y_i) - KL(Pz_i^{m+1}, Pz_i^m)\}$ goes to zero, so $\{KL(z^{m+1}, z^m)\}$ goes to zero. If z^* is any cluster point of $\{z^m\}$ then z^* is left unchanged by the iterative process, so $y = Pz^*$. Using $x = z^*$ in MART 3, we find that $\{KL(z^*, z^m)\}$ is decreasing; it follows that $\{z^m\}$ converges to z^*. The difference $KL(x, z^m) - KL(x, z^{m+1})$ is independent of x, so long as $y = Px$; therefore, $KL(x, z^0) - KL(x, z^*)$ is also independent of x. It follows that minimizing $KL(x, z^0)$ over all $x \geq 0$ for which $y = Px$ is equivalent to minimizing $KL(x, z^*)$ over the same x. But the solution to the latter problem is obviously $x = z^*$. The theorem is proved. □

In [18] Censor and Segman generalize the MART algorithm to include block-iterative versions, in which disjoint subsets of the equations are used at each step. MART is an extreme case, in which each subset corresponds to a single equation. At the other extreme, they use all the equations within each step; their algorithm is similar to the SMART algorithm, but includes a weighting term. Their proof of convergence seems to require the weighting. For the SMART case our normalizing the column sums to one eliminates the need for the weighting; for MART the normalization is not required. They also do not consider non-constant z^0. In [19] it is shown that block-iterative versions of MART and SMART converge in the consistent case, for all partitions of the equations into blocks, but that no such general result is available for EMML.

In the inconsistent case, in which $y = Px$ has no nonnegative solutions, the behavior of MART remains an open question. In every simulation we have considered, MART produces a limiting cycle of vectors, one for each equation, generally; in this respect, the behavior of MART is analogous to that of ART in the inconsistent case. Proving the existence of the limit cycle is complicated by the fact that, unlike ART, the MART is not a (KL) contraction, in general, and unlike SMART, we have no characterization of the vectors of the limit cycle as the solution of some optimization. On the positive side, we can prove the MART version of EMML 11 and SMART 11, which shows that the vectors of the limit cycle, when they exist, have no more than I-1 nonzero entries. Because the vectors of the limit cycle are not actually on the hyperplanes we can use them to extract a new "data" vector, to replace y. If we then repeat the MART, with this new y replacing the previous one and with the supports restricted to that of the previous limit cycle vectors, we obtain a nested iterative scheme that eventually converges to a singleton limit cycle; this singleton can be characterized as the limit of the SMART algorithm, using the original y but subject to the added constraint that the support be limited to that of the singleton. This result assumes, of course, that the limiting cycles exist at each step of the way.

A special case of the MART algorithm (3.2), in which the matrix P has entries that are either zero or one, was considered in the 1930's by Sheleikhovskii and by Kruithof. In [20] Bregman proves MART 4 for this special case and extends these ideas to include more general nonlinear projections in [21]. In [22] Krupp discusses Kruithof's method and considers the case of general nonnegative matrix P. Krupp presents the form of the solution for general P, but does not explicitly describe the MART algorithm for this general case; some have suggested that he discusses the SMART algorithm, but a close reading of [22] shows that not to be the case. The MART algorithm as presented in [15] is also this special case of (9), but in [23] the full MART algorithm is considered.

4. Acknowledgements. This work was supported by grants CA42165 and CA51071 from the National Cancer Institute, with principal investigators Mike King and Bill Penney of the Department of Nuclear Medicine, University of Massachusetts Medical Center, Worcester, Massachusetts. Its contents are soley the responsibility of the author and do not necessarily represent the official views of the National Cancer Institute. The author wishes to thank Professor Yair Censor of the University of Haifa for helpful

discussions on these matters.

REFERENCES

[1] A. Rockmore and A. Macovski, *A maximum likelihood approach to emission image reconstruction from projections*, IEEE Trans. Nucl. Sci., Vol. NS-23, 1976, pp. 1428–1432.

[2] L. Shepp and Y. Vardi, *Maximum likelihood reconstruction for emission tomography*, IEEE Trans. Med. Imag., Vol. MI-1 (2) Oct. 1982, pp. 113–122.

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

[4] K. Lange and R. Carson, *EM reconstruction for emission and transmission tomography*, J. Comput. Assist. Tomog., Vol. 8, 1984, pp. 306–312.

[5] Y. Vardi, L. Shepp and L. Kaufman, *A statistical model for positron emission tomography*, J. Amer. Statist. Assoc., Vol. 80, 1985, pp. 8–37.

[6] I. Csiszar and G. Tusnady, *Information geometry and alternating minimization procedures*, Statistics and Decisions, suppl. issue No. 1 (1984), pp. 205–237.

[7] K. Lange, M. Bahn and R. Little, *A theoretical study of some maximum likelihood algorithms for emission and transmission tomography*, IEEE Trans. Med. Imag., Vol. MI-6(2), June 1987, pp. 106–114.

[8] D. Titterington, *On the iterative image space reconstruction algorithm for ECT*, IEEE Trans. Medical Imaging, Vol. MI-6, pp. 52–56, Mar. 1987

[9] P. Schmidlin, *Iterative separation of sections in tomographic scintigrams*, Nuclear Medicine, Schatten Verlag, Stuttgart, Vol. 15, No. 1, 1972.

[10] S. Holte, P. Schmidlin, A. Linden, G. Rosenqvist and L. Eriksson, *Iterative image reconstruction for positron emission tomography: a study of convergence and quantitation problems*, IEEE Trans. Nuclear Science, Vol. 37, No. 2, April 1990, pp. 629–635.

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

[12] I. Csiszar, *A geometric interpretation of Darroch and Ratcliff's generalized iterative scaling*, Annals of Statistics, Vol. 17, No. 3, 1989, pp. 1409–1413.

[13] C. Byrne, *Iterative image reconstruction algorithms based on cross-entropy minimization*, IEEE Trans. Image Processing, Vol. 2, No. 1, Jan. 1993, pp. 96–103.

[14] C. Byrne, *Erratum and addendum to 'Iterative image reconstruction algorithms based on cross-entropy minimization'*, to appear in IEEE Trans. Image Proc., Feb. 1995.

[15] R. Gordon, R. Bender and G. Herman, *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography*, J. Theoret. Biol., Vol. 29 (1970), pp. 471–481.

[16] A. Fiacco and G. McCormick, *Nonlinear Programming: sequential unconstrained minimization techniques*, SIAM reissue (1990).

[17] A. Lent, *A convergent algorithm for maximum entropy image restoration with a medical x-ray application*, in *Image Analysis and Evaluation* (R. Shaw, ed.), SPSE, Wash. D.C. (1977), pp. 238–243.

[18] Y. Censor and J. Segman, *On block-iterative entropy maximization*, J. Inform. and Optim. Sciences, Vol. 8(3) (1987), pp. 275–291.

[19] C. Byrne, *Block-iterative methods for image reconstruction from projections: some convergence results*, submitted to IEEE Trans. Image Proc., Dec. 1993.

[20] L. Bregman, *Proof of the convergence of Sheleikhovskii's method for a problem with transportation constraints*, U.S.S.R. Comput. Math. and Math. Phys., Vol. 7(1) (1967), pp. 147–156.

[21] L. 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. Comput. Math. and Math. Phys., Vol. 7(3) (1967), pp. 200–217.

[22] R. Krupp, *Properties of Kruithof's projection method*, Bell System Tech. Jour., Vol. 58(2) (Feb. 1979), pp. 517–538.

[23] G. Herman and A. Lent, *Iterative reconstruction algorithms*, Comput. Biol. Med., Vol. 6 (1976), Pergamon Press, pp. 273–294.