Convergent Block-Iterative Algorithms for Image Reconstruction from Inconsistent Data

Charles L. Byrne, *Member, IEEE*

***Abstract*— It has been shown recently that convergence to a solution can be significantly accelerated for a number of iterative image reconstruction algorithms, including simultaneous Cimmino-type algorithms, the "expectation maximization" method for maximizing likelihood (EMML) and the simultaneous multiplicative algebraic reconstruction technique (SMART), through the use of rescaled block-iterative (BI) methods. These BI methods involve partitioning the data into disjoint subsets and using only one subset at each step of the iteration. One drawback of these methods is their failure to converge to an approximate solution in the inconsistent case, in which no image consistent with the data exists; they are always observed to produce limit cycles (LC's) of distinct images, through which the algorithm cycles. No one of these images provides a suitable solution, in general. The question that arises then is whether or not these LC vectors retain sufficient information to construct from them a suitable approximate solution; we show here that they do. To demonstrate that, we employ a "feedback" technique in which the LC vectors are used to produce a new "data" vector, and the algorithm restarted. Convergence of this nested iterative scheme to an approximate solution is then proven. Preliminary work also suggests that this feedback method may be incorporated in a practical reconstruction method.**

***Index Terms*—Restoration, tomography.**

I. Introduction and Background

Many image reconstruction methods involve the exact or approximate solution of large systems of equations, usually with side constraints, such as nonnegativity or local smoothness. Those methods based on optimizing some objective function, such as likelihood, can often be reformulated in terms of solving such a system. Because of the many unknowns and data values involved, and because of the relative ease with which side constraints can be included, iterative techniques are increasingly attractive. Some of the most popular iterative techniques, such as the "expectation maximization" approach to maximizing likelihood (EMML), as used in single photon emission computed tomography (SPECT) and positron emission tomograpy (PET) converge too slowly, however. It has been shown recently [1] that EMML and related algorithms, such as the simultaneous multiplicative algebraic reconstruction technique (SMART), can be significantly accelerated through the use of the block-iterative (BI) approach and block-dependent equation rescaling; similar results for the algebraic reconstruction technique (ART) have been obtained by Herman and Meyer [2] and Guan and Gordon [3]. In BI methods, one begins by partitioning the data into disjoint subsets; each step of the iteration involves only one of these subsets. If there is but one subset, containing all the data, the method is said to be "simultaneous." The "rescaled" versions of these BI methods, RBI-EMML and RBI-SMART, converge rapidly to a solution, whenever it is possible to fit an image exactly to the data; this is the so-called feasible case. One drawback to the use of BI methods is their behavior in the inconsistent case, in which no image consistent with the data exists.

When no image is consistent with the data, we typically seek an approximate solution, such as a least squares estimate. The BI methods do not produce such an approximate solution, however; they produce a limit cycle (LC) of distinct images, as many as there are blocks, through which the iteration cycles. How distinct these images are depends on how noisy the data is. No one of these images provides a suitable reconstruction, in general. The obvious question is whether or not the LC retains sufficient information to construct from it a suitable approximate solution; we show here that, for ART, BI-EMML, and BI-SMART, the answer is that it does. One might hope that there would be a simple calculation that could be performed on the vectors of the LC to produce an approximate solution; for example, a weighted mean of some sort. Although that remains an attractive possibility, we have not been able to obtain such a result. We introduce a "feedback" approach, in which new "data" is extracted from the vectors of the LC and the algorithm restarted. This nested iterative scheme is shown to converge to an approximate solution in each case.

Our main goal in this paper is to provide a theoretical justification for the use of BI methods in the inconsistent case. The feedback approach is used here as a tool to construct the theory. Whether or not the feedback can be made part of a practical image reconstruction procedure remains to be demonstrated, although successful application of feedback on small-scale problems has been encouraging. In actual applications the inner iterative loop, involving the BI algorithm, would probably include a regularizing term to penalize roughness. This still leads to an LC, however. The outer loop, the feedback step, would have to be performed only a small number of times if the acceleration advantage of the BI method is not to be lost.

Manuscript received November 27, 1995; revised October 16, 1996. This work was supported in part by Grants CA42165 and CA51071 from the National Cancer Institute. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Andrew Yagle.

The author is with the Department of Mathematical Sciences, University of Massachusetts at Lowell, Lowell, MA 01854 USA (e-mail: byrnec@woods.uml.edu).

Publisher Item Identifier S 1057-7149(97)06249-0.

II. Overview

The algebraic reconstruction technique (ART) and its multiplicative version (MART) [4], are extreme examples of BI methods; each block consists of a single data value. Both algorithms converge to a solution whenever possible, but if they cannot, they produce limit cycles. The existence of the LC for ART was demonstrated by Tanabe [5]; no proof of existence of LC is known for any of the BI methods, including MART, although we always see them.

When the M by N system of linear equations $Ax = b$ is overdetermined ($M > N$) it typically has no solution; in that case the ART does not converge to a single vector, but to a limit cycle (LC) of (usually) M distinct vectors, say $\{z^1, \cdots, z^M = z^0\}$ [5]. Denote by Az_m the mth entry of the vector Az. For c the vector with entries $c_m = Az_m^{m-1} = (Az^{m-1})_m$, the systems $Ax = b$ and $Ax = c$ have the same (assume unique) least squares solution; call it x^*. With $b^0 = b$, $b^1 = c$, and in general $b^{k+1} = c^k$, let $\mathrm{LC}(k) = \{z^{k,1}, \cdots, z^{k,M} = z^{k,0}\}$ be the limit cycle obtained by applying ART to the system $Ax = b^k$, and $c^k = (Az_m^{k,m-1})$. We find that for each m the sequence of vectors $\{K^{-1}(z^{0,m} + z^{1,m} + \cdots + z^{K-1,m})\}$ converges to x^*, as K goes to infinity. Also the sequence $\{K^{-1}(b^0 + b^1 + \cdots + b^{K-1})\}$ converges to Ax^*.

For the multiplicative ART (MART) the situation is both simpler and more complicated. First, MART applies to systems of I equations in J unknowns $y = Px$ where $y > 0$, $P \geq 0$ and a nonnegative x is sought. For convergence in the feasible case the entries of P must be in [0, 1]. When there is no nonnegative solution of $y = Px$ MART behaves like ART: it gives an LC denoted by $\{z^1, \cdots, z^I = z^0\}$; no proof of this is known, but it is always observed. Defining the vector $y^1 = (Pz_i^{i-1})$ and applying MART to the new system $y^1 = Px$, we get a new LC, just as in ART. The sequence $\{y^k\}$ obtained in this way converges to a vector y^∞ for which the system $y^\infty = Px$ has a nonnegative solution. If $J < I$ and all the LC vectors are strictly positive then the solution of $y^\infty = Px$ is the unique minimizer of the Kullback–Leibler or cross-entropy distance $\mathrm{KL}(Px, y)$ over all $x \geq 0$.

In Section III, we discuss the ART algorithm. Its simultaneous version, due to Trummer [6], converges, in the inconsistent case, to the least squares solution of $Ax = b$ closest to the starting vector, in the Euclidean distance. Restricting ourselves to the case of a unique least squares solution, we investigate the relationship between that solution and the LC produced by ART. We find that, in the special case in which M, the number of equations, is one more than N, the number of unknowns, all the vectors of the LC lie on a sphere in N-space having the least squares solution at the center; for other choices of M and N this is no longer true. This result suggests that the least squares solution can be obtained just from the vectors of the LC. Using feedback, we show that this is generally the case.

In Section IV, we examine the BI-SMART, with MART as a special case. The simultaneous version, SMART, produces a nonnegative approximate solution in the inconsistent case. It has been shown [7]–[9], that when there are more unknowns (J) than equations (I), this approximate solution has at most $I - 1$ nonzero entries. Limiting ourselves to the case of $J < I$ and using feedback, we show that we can recover the SMART estimate from the LC.

We take up the recently discovered BI-EMML [1], [10] in Section V. In simulations, we find the feedback approach successful in obtaining an approximate solution, although we have proven convergence for only a special case.

We include proofs of all our results in the Appendices. If there are methods more efficient than the feedback approach, one way to find them might be to examine the role played by the feedback in the derivation of the theory.

III. The ART Algorithm

The ART algorithm [4], also called Kaczmarz's method [11], is an iterative procedure for solving a general linear system of equations, $Ax = b$. The algorithm consists of projecting the current vector orthogonally onto the hyperplane determined by the current equation and proceeding through the equations one at a time (usually cyclically) until convergence. When the system has no solution the ART produces a limit cycle of vectors, typically one for each equation, rather than a single vector.

We show that the limit cycle can be used to extract a new "right-hand side" or b vector. When the algorithm is restarted using this new b a new limit cycle is reached. Proceeding iteratively in this way, we get a sequence of vectors $\{b^k\}$, whose sequence of successive averages converges to a vector d for which the exact solution to $Ax = d$ is the (normalized) least squares solution of the original problem.

We consider only a fixed (but random) ordering of the equations. For the M by N system of equations $Ax = b$, the projection of vector z onto the mth hyperplane $B_m = \{x | Ax_m = b_m\}$ is denoted by $P_m(z)$ and its nth entry is given by

$$P_m(z)_n = z_n + \frac{A_{m,n}(b_m - Az_m)}{\sum_n A_{m,n}^2} \tag{1}$$

where $\sum_n$ denotes the sum over index n. For notational convenience, we assume that the original equations have been rescaled so that $\sum_n A_{m,n}^2 = 1$ for all m; then the denominator in (1) disappears.

We can introduce relaxation in ART by taking the successor z' to z to be not $P_m(z)$ but the vector

$$z'_n = z_n + tA_{m,n}(b_m - Az_m) \tag{2}$$

where, for now, t is any quantity (possibly depending on m or on the actual iteration number). Then the L^2 distance between z and its successor is

$$L^2(z', z) = t^2 L^2(b_m, Az_m) \tag{3}$$

here $L^2(a, b) = \sum_n (a_n - b_n)^2$ for any vectors $a = (a_n)$ and $b = (b_n)$ of any length. The (relaxed) ART algorithm begins with any z^0 and, for $k = 0, 1, \cdots$ and $m = k(\mathrm{mod}\, M) + 1$, sets $z^{k+1} = (z^k)'$ as in (2).

Kaczmarz [11] proved that, provided A is square and invertible, in which case there is a unique solution, (1) converges to that solution. In [12], Herman *et al.* showed that, with restriction on the t, relaxed ART converges to a solution whenever there is one; Trummer [13] weakened the restriction on the t somewhat. The case in which no solution exists has been considered by Tanabe [5], who proves convergence to a limit cycle of vectors, by Eggermont *et al.* [14], by Dax [15], and by Censor *et al.* [16].

Extending the Tanabe result to the relaxed case, with $M > N$ and A full rank, we have the following (see Appendix A for the proof).

Theorem 1: For t in (0, 2), the relaxed ART method (2) exhibits cyclic convergence, that is, for each fixed m, the subsequences $\{z^{jM+m}\}$ converge to distinct limits, denoted $z^{\infty, m}$. The LC is then $\{z^{\infty, 1}, z^{\infty, 2}, \cdots, z^{\infty, M} = z^{\infty, 0}\}$.

For notational simplicity, we now write $z^{\infty, m} = z^m$, $m = 0, 1, \cdots, M-1$; the limit cycle LC is then $\{z^m, m = 0, 1, 2, \cdots, M-1\}$. The size of the LC is related to the extent to which a solution is not available; specifically, we have from (3) that

$$\sum_m L^2(z^m, z^{m-1}) = t^2 \sum_m L^2(b_m, Az_m^{m-1}). \quad (4)$$

We shall assume, throughout this section, that the matrix A has full rank, so that the least squares solution of $Ax = b$, x^*, is unique. If, as we let the relaxation parameter t approach zero, the elements of the LC approach individual limits, these must all be the same vector. This has been considered by Censor *et al.* [16]. They show that the limiting single vector is, in fact, the least squares solution of the system $Ax = b$. Specifically, we have the following result, proven using (4) in Appendix A.

Theorem 2 [16]: let $\{t_k\}$ be a sequence of positive quantities converging to zero. For each k let $\text{LC}(k) = \{z^{k,0}, \cdots, z^{k, M-1}, z^{k, M} = z^{k, 0}\}$ be the ART limit cycle obtained using $t = t_k$ in (2). As $\{t_k\}$ converges to zero, the vectors in LC(k) converge to x^*.

The difficulty with using this approach to obtain the least squares solution is that as the relaxation parameter goes to zero the convergence slows. We want a different way to extract the least squares solution from the LC.

In the special case of $M = N+1$ and $t = 1$ there is a curious relationship between the elements $\{z^m\}$ of the LC and the least squares solution x^*. We have the following theorem, proven in Appendix A.

Theorem 3: For $M = N+1$ and $t = 1$, the elements of the LC lie on a sphere in N-dimensional space having the least squares solution x^* at the center.

From now on we shall consider only $t = 1$ in (2).

This theorem suggests again that there is some relationship between the LC and the least squares solution, but in general the vectors of the ART LC do not lie on a sphere. We have not found any way to calculate the least squares solution simply in terms of the vectors of a single LC. For $c = (Az_m^{m-1})$, the systems $Ax = b$ and $Ax = c$ have the same least squares solution; this follows from $z_n^m - z_n^{m-1} = A_{m,n}(b_m - Az_m^{m-1})$ by summing on index m on both sides. This suggests the following feedback method: With $b^0 = b$, $b^1 = c$, and in general $b^{k+1} = c^k$, let $\text{LC}(k) = \{z^{k,1}, \cdots, z^{k,M} = z^{k,0}\}$ be the limit cycle obtained by applying ART to the system $Ax = b^k$, and $c^k = (Az_m^{k, m-1})$. We shall prove the following (see Appendix A).

Theorem 4: For each m, the sequence $\{K^{-1}(z^{0,m} + z^{1,m} + \cdots + z^{K-1, m})\}$ converges to the least squares solution x^* of $Ax = b$. Also, the sequence $\{K^{-1}(b^0 + b^1 + \cdots + b^{K-1})\}$ converges to Ax^*.

We see from the theorem how, by extracting a new b from the LC at each step, we can form a new sequence that converges to a single vector, namely the least squares solution.

The ART does not involve a nonnegativity constraint. We can, of course, seek to minimize $L^2(b, Ax)$ over $x \geq 0$. When there is no exact nonnegative solution for $Ax = b$ these nonnegative least squares solutions must have no more than $M-1$ positive entries. In many image processing applications the quantity N is the number of pixels chosen and is often large, to permit high resolution reconstruction. This result, our Theorem 5 below, proven in Appendix A, is sometimes called a "night sky" theorem because it asserts that the resulting optimal image x will have a few nonblack pixels against a background of many black pixels. We have analogous results for SMART, EMML, and their BI versions (see Theorem 6 in Section IV and [7, Props. 1 and 2]).

Theorem 5: Suppose that A and every submatrix obtained from A by deleting columns has full rank. Suppose also that the system $Ax = b$ has no nonnegative solution. There is a subset S of $\{1, 2, \cdots, N\}$ with cardinality at most $M-1$, such that, for all $z = x \geq 0$ solving the problem of minimizing $L^2(b, Az)$, subject to $z \geq 0$, $x_n > 0$ only if n is in S. Therefore, x is unique.

IV. The BI-SMART Algorithm

A number of iterative algorithms for the reconstruction of images from projections can be obtained by alternately minimizing certain distances between convex sets. Of particular interest here is the so-called cross-entropy or Kullback–Leibler (KL) distance [17] between two nonnegative vectors $a = (a_n)$ and $b = (b_n)$, defined to be $\text{KL}(a, b) = \sum_n a_n \log(a_n/b_n) + b_n - a_n$, with the understanding that $0 \log 0 = 0$ and that $\text{KL}(a, b) = +\infty$ if there is an n with $a_n > 0$ and $b_n = 0$. In previous articles [7]–[9], it was shown that by using KL distances in various ways, one can obtain the MART, a variant of MART, the SMART, the EMML algorithm for likelihood maximization in emission tomography, and regularized versions of the latter two methods. Each of these algorithms can be viewed as providing approximate nonnegative solutions to a (possibly inconsistent) linear system of equations, $y = Px$, where $y = (y_i) \geq 0$, $P = (P_{i,j})$, $P_{i,j} \geq 0$ for all $i = 1, \cdots, I$, $j = 1, \cdots, J$, and we assume the problem has been normalized so that the column sums of P equal one. Proofs of results in this section are in Appendix B.

The EMML and SMART algorithms always converge (see e.g., [7]–[9]), and BI-SMART (and, therefore, MART) converges, provided the system $y = Px$ is consistent; that is, there is a nonnegative x for which $y = Px$ [9], [10], [18]. The behavior of BI-SMART in the inconsistent case is an open question. In the inconsistent case BI-SMART fails to converge. However, in every example we have considered, the subsequences of the full BI-SMART sequence associated with completed cycles do converge, but to distinct limits; we shall refer to these vectors as the limiting cycle of BI-SMART, analogous to what we just saw with ART. In contrast, the SMART algorithm converges in the inconsistent case to the unique nonnegative minimizer of $\mathrm{KL}(Px, y)$ for which the distance $\mathrm{KL}(x, x^0)$ is minimized. We shall assume convergence of BI-SMART to a limiting cycle and consider the consequences of feedback, which can be viewed as a particular form of relaxation or hyperplane modification.

In the inconsistent case both the SMART limit and the limiting cycle vectors of BI-SMART have support (indices of nonzero entries) with cardinality less than the number of equations, although there need not be any relationship between the two support sets [7]–[9]. When the BI-SMART subsequences have converged, we can extract the new projection data and reapply BI-SMART, with this new data replacing the original vector y. We can limit the reconstruction to the support set S, or not, as we wish. Restarting BI-SMART with this new y vector we obtain a new limit cycle and a third set of projections, not usually the ones fed back in. This suggests that the usual BI-SMART algorithm is only part of a more complete algorithm, involving repeated feeding back of the limit cycle projection data until convergence. We prove here that, assuming convergence of BI-SMART to a limit cycle, the expanded version of BI-SMART in which the support is also used converges to the unique nonnegative vector x minimizing $\mathrm{KL}(Px, y)$ and supported on the (nonempty) intersection S' of the supports of the limit cycles. If the support of the SMART limit is contained within S' then the limit of the expanded BI-SMART is that of SMART. The expanded BI-SMART that does not use the limit cycle support also converges to a single vector, but we do not as yet have a description of what it is.

In [19], DePierro and Iusem consider relaxation for Bregman's method. They show that the appropriate way to relax these nonlinear methods is to modify the y vector prior to Bregman projection. For linear orthogonal projection, as used in ART [4], the modification of the y (moving the hyperplanes) is equivalent to inserting a multiplicative factor in the increment, as in (2), so relaxation in ART has focused on the modification of that factor (e.g., see Censor, *et al.* [16]). The expanded BI-SMART considered here is related to DePierro–Iusem relaxation, in which the modification of the y vector (equivalently, the amount of relaxation) is determined automatically by the previous BI-SMART limit cycle.

The regularized SMART algorithm [7]–[9] was designed to minimize the functional $G(x) = a\mathrm{KL}(Px, y) + (1 - a)\mathrm{KL}(x, p)$, where $0 \le a \le 1$ and $p \ge 0$ is a prior estimate of the desired x. Starting with $x^0 > 0$ we define, for each $j = 1, \cdots, J$

$$x_j^{m+1} = (x_j^m)^a (p_j)^{1-a} \exp\left[a \sum_i P_{i,j} \log\left(\frac{y_i}{Px_i^m}\right)\right] \tag{5}$$

with the sum over $i = 1, \cdots, I$. Equivalently, we can write

$$x_j^{m+1} = (x_j^m)^a (p_j)^{1-a} \prod_i \left(\frac{y_i}{Px_i^m}\right)^{aP_{i,j}} \tag{6}$$

with the product over $i = 1, \cdots, I$. For $a = 1$ this is closely related to the MART of Gordon *et al.* [4].

Lent [18] has shown that if x^0 is constant and $y = Px$ has nonnegative solutions then the MART iteration converges to the solution that maximizes the Shannon entropy. In [20], Censor presents a regularized version of MART and remarks that the behavior of MART in the inconsistent case is unknown. In [21], Herman notes that in the inconsistent case MART may not maximize entropy. As we showed in [7], the SMART iteration scheme (5) converges to the unique minimizer of $G(x)$ for $a < 1$. For $a = 1$ SMART converges to the unique nonnegative solution of $y = Px$ minimizing $\mathrm{KL}(x, x^0)$, in the consistent case. In the inconsistent case, SMART converges to the unique nonnegative minimizer of $\mathrm{KL}(Px, y)$ minimizing $\mathrm{KL}(x, x^0)$; the support has cardinality less than the number of equations, for almost all P, so the minimizer of $\mathrm{KL}(Px, y)$ is almost always unique in the inconsistent case.

The (unrelaxed) MART algorithm of Gordon *et al.* [4] is the following: Starting with $z^0 = x^0 > 0$, let

$$z_j^{k+1} = z_j^k \left(\frac{y_i}{Pz_i^k}\right)^{P_{i,j}} \tag{7}$$

for $k = 0, 1, 2, \cdots$ and $i = i(k) = k(\mathrm{mod}\, I) + 1$; let $x^n = z^{nI}$, $n = 0, 1, 2, \cdots$, where $i = 1, \cdots, I$. We then have

$$x_j^{n+1} = x_j^n \prod_i \left(\frac{y_i}{Pz_i^{nI+i-1}}\right)^{P_{i,j}} \tag{8}$$

with the product over i.

Compare MART with the SMART iteration (for $a = 1$) defined by

$$x_j^{m+1} = x_j^m \prod_i \left(\frac{y_i}{Px_i^m}\right)^{P_{i,j}}. \tag{9}$$

Because in MART, the factors in the product depend on the most recent z^{nI+i-1}, while in (9) the factors are computed simultaneously, using the current x^m, we call (9) the "simultaneous" MART (SMART) algorithm.

The MART converges in the consistent case; the sequence of iterates beginning at $x^0 > 0$ converges to the unique nonnegative solution of $y = Px$ for which $\mathrm{KL}(x, x^0)$ is minimized [9], [10]; when x^0 is a constant vector the limit is the maximum Shannon entropy solution, as Lent [18] has shown. In the inconsistent case, the sequence of iterates cannot converge without violating the assumption of inconsistency.

In [1], we presented a BI version of SMART, called BI-SMART. The set $\{i = 1, 2, \cdots, I\}$ is partitioned into disjoint subsets $S_1, \cdots, S_N$. Then, for $n = k(\mathrm{mod}\, N) + 1$, $k = 0, 1, \cdots$, the BI-SMART iteration scheme is

$$z_j^{k+1} = z_j^k {\prod}^n \left(\frac{y_i}{Px_i^k}\right)^{P_{i,j}} \tag{10}$$

where $\prod^n$ denotes the product over those indices i in S_n. In [1], the BI algorithms involved a block-dependent rescaling of the equations to accelerate convergence. In this paper, we shall assume, without changing notation, that for each n and each i in S_n the ith equation in $y = Px$ has been divided by the rescaling factor $m_n = \max_j \{\sum^n P_{i,j}\}$ where $\sum^n$ denotes the sum over those indices i in S_n. The BI algorithms we discuss here are then automatically the "rescaled" accelerated versions, called the RBI methods in [10].

Comparison with the SMART iteration suggests that perhaps for each fixed n the subsequence $\{z^{mN+n}, n \text{ fixed}, m = 0, 1, \cdots\}$ obtained from the BI-SMART sequence by selecting every Nth term might converge. In fact, in every computed example we have considered, this has been the case; the sequences $\{z^{mN+n}\}$ do converge, but to distinct limits $z^{\infty,n}$, which we shall call the limit cycle. As with the SMART algorithm, the vectors of the limit cycle have common support S with cardinality less that the number of equations (see Theorem 6). For each n and index i in S_n we take the new projection data $\{Pz_i^{\infty,n-1}\}$ in place of the $\{y_i\}$ and repeat the BI-SMART iteration, restricted to S or not, until the subsequences have again converged, we find that we have a new set of projection data, not the $\{Pz_i^{\infty,n-1}\}$ obtained the previous time. This is the feedback procedure mentioned above.

The BI-SMART algorithm is best analyzed in terms of alternating minimization of certain distances, just as was done in [7]–[9] for the SMART, EMML, and MART. For each $x \geq 0$ and $z \geq 0$ and for each $n = 1, \cdots, N$ let

$$G_n(x, z) = \text{KL}(x, z) - \sum^n \text{KL}(Px_i, Pz_i) + \sum^n \text{KL}(Px_i, y_i). \tag{11}$$

Since the columns of P sum to one it follows [7]–[9] that $\text{KL}(x, z) \geq \text{KL}(Px, Pz)$; this is true because $\text{KL}(a, b) = \text{KL}(a_+, b_+) + \text{KL}[a, (a_+/b_+)b]$, where $a_+ = \sum_n a_n$, $b_+ = \sum_n b_n$. So we have $G_n(x, z) \geq 0$. Then, a direct calculation shows that, for any x and z and for any n we have

$$G_n(x, z) = G_n(z', z) + \text{KL}(x, z') \tag{12}$$

where, for specified n, we define $z'_j = z_j \prod^n (y_i/Pz_i)^{P_{i,j}}$, $j = 1, \cdots, J$. For any x and z and for any n the nonnegative quantity $G_n(x, z)$ is minimized with respect to z by the choice $z = x$. In fact, we have $G_n(x, z) = G_n(x, x) + \text{KL}(x, z) - \sum^n \text{KL}(Px_i, Pz_i)$, and $G_n(x, x) = \sum^n \text{KL}(Px_i, y_i)$. We show that the sequence of iterates $\{z^k\}$ is contained within a bounded set, so that it has subsequential limit points.

To that end, for each j, let $M_j = \text{maximum}\{y_i/P_{i,j}, i = 1, 2, \cdots, I\} < +\infty$, with the maximum taken over only those i for which $P_{i,j} > 0$, and let $C_j = \text{maximum}\{z_j^0, M_j\}$.

Lemma 1: For each k and each j we have $z_j^k \leq C_j$.

In discussions of BI-SMART it is always sensible to assume that if any row of P is all zero then it is disregarded. Similarly, if any $y_i = 0$ that equation is disregarded, since including that term in the formula (10) produces a zero estimate.

We shall assume from now on that BI-SMART always converges to a limit cycle whenever the system of equations has no nonnegative solution. Then, for each fixed n and i in S_n, the sequence $\{Pz_i^{mN+n-1}\}$ converges to $w = \{w_i = Pz_i^{\infty,n-1}\}$. In the inconsistent case, $\sum_i P_{i,j} \log(y_i/w_i) = 0$ for fewer j than the number of equations, so that the vectors of the limit cycle have support smaller than the number of equations. To prove this, we need to assume that P has the full rank property (FRP): P and all matrices obtained from P by deletion of columns have full rank.

Theorem 6: Let S be the support set for the vectors of the limit cycle. Then, if $y = Px$ has no nonnegative solution and if P has the FRP, S has fewer members than there are equations.

Lemma 2: For all $j = 1, \cdots, J$, $\sum_i P_{i,j} \log(y_i/w_i) \leq 0$, with equality for j in S.

Corollary: For all $x \geq 0$ we have

$$\text{KL}(Px, y) \geq \text{KL}(Px, w) + \sum_i y_i - \sum_i w_i \tag{13}$$

with equality if the support of x is within S.

The feedback procedure is to set $y^0 = y$, $w^0 = (Pz_i^{\infty,n-1})$, where n is such that i is in S_n, and for $k = 1, 2, \cdots$ set $y^k = w^{k-1}$; then restart the BI-SMART iteration to obtain the kth limit cycle, $\text{LC}(k) = \{z^{k,0}, \cdots, z^{k,N-1}, z^{k,N} = z^{k,0}\}$.

Lemma 3: The sequence $\{\sum_i y_i^k\}$ is decreasing, so that $\{\text{KL}(Px, y^k)\}$ is decreasing, for all $x \geq 0$, and the sequence $\{\sum_n G_n(z^{k,n}, z^{k,n-1})\}$ converges to zero.

Proof: We have $\sum_i y_i^k - \sum_i y_i^{k+1} = \sum_n G_n(z^{k,n}, z^{k,n-1}) \geq 0$. Now use the Corollary to Lemma 2.

We can extend the boundedness Lemma 1 to show that the set $\{z^{k,n}, n = 1, \cdots, N, k = 0, 1, \cdots\}$ is bounded. From the sequence $\{z^{k,0}\}$ we extract a subsequence $\{z^{k_m,0}\}$ converging to $z^{*,0}$. The sequence of BI-SMART successors $\{z^{k_m,1}\}$ then has subsequence converging to some $z^{*,1}$. Continuing in this manner, we obtain $\{z^{*,0}, \cdots, z^{*,N-1}\}$. With minor restrictions on the matrix P, we show that $z^{*,0} = z^{*,1} = \cdots = z^{*,N-1}$. To avoid pathological special cases in the proof of the theorem below, we make the technical assumption that for every n there is a j with $z_j^{*,n} > 0$ and $0 < P_{i,j} < 1$ for some i in S_n. We then have the following theorem.

Theorem 7: The sequence $\{y^k\}$ converges to a vector y^∞ for which the linear system of equations $y^\infty = Px$ has an exact nonnegative solution. If, in addition, $J < I$ and every $\text{LC}(k)$ has strictly positive vectors, then the solution of $y^\infty = Px$ minimizes $\text{KL}(Px, y)$ over all $x \geq 0$; therefore, it is the same limit obtained by the SMART algorithm.

We obtain a similar result for general I and J, without requiring that the vectors of the $\text{LC}(k)$ be strictly positive, if, at each restarting of the BI-SMART we restrict j to the support of the most recently obtained LC.

We sketch the proof of Theorem 7 in Appendix B. This result can be viewed as the analog for BI-SMART of the theorem of Censor *et al.* [16] to the effect that when underrelaxed ART is repeated with relaxation constant going to zero the points of the ART limit cycle converge to the least squares solution of the original (normalized) system of equations. Indeed, from the work of DePierro and Iusem [19], we know that relaxation in Bregman iteration should be viewed as modification of the hyperplanes prior to nonlinear projection. Here we modify the hyperplanes when we replace y^{k-1} with $y^k = w^{k-1}$. The

x' minimizing KL(Px, y) can be viewed as a generalized solution of the system, analogous to finding the least squares "solution" by minimizing the Euclidean distance between y and Px.

V. The BI-EMML Algorithm

In [1], the EMML algorithm was extended to a BI version, called the BI-EMML. This BI algorithm converges, in the consistent case, for any configuration of subsets, and its "rescaled" version reduces to the "ordered subset" method (OS-EM) of Hudson *et al.* [23] when their "subset balance" condition applies. The BI-EMML algorithm is the following, again assuming the rescaling has been done: beginning with $z^0 > 0$, and with $n = k(\text{mod}\, N) + 1$, define

$$z_j^{k+1} = (1 - \sum^n P_{i,j})z_j^k + z_j^k \sum^n P_{ij} \frac{y_i}{Pz_i^k}. \tag{14}$$

For the case in which each set S_n is a singleton, (14) gives the EMML version of MART, which we call EMART. In the inconsistent case, this algorithm is always observed to produce a LC, although no proof of this is yet known. We apply feedback to the BI-EMML, just as to the BI-SMART.

Let $y^0 = y$ and let LC(0) $= \{z^{0,1}, \cdots, z^{0,N} = z^{0,0}\}$ be the limit cycle obtained using BI-EMML on y^0. Using LC(0), we set $w_i^0 = Pz_i^{0,n-1}$ for each i, where n is chosen so that i is in S_n. Then let $y^1 = w^0$ and restart BI-EMML using data y^1. In this way, we generate a sequence $\{y^k\}$ and corresponding limit cycles LC(k) $= \{z^{k,1}, \cdots, z^{k,N} = z^{k,0}\}$. The conjecture is: i) $\{y^k\}$ converges always to a limit vector, say y^∞; ii) the system $y^\infty = Px$ has a nonnegative solution, say x^∞; and iii) the limit cycles LC(k) converge to x^∞.

There is proof for i) and ii) if there are at least as many pixels as data values, so P has at least as many columns as rows (and is assumed full rank); and for iii) if P is square and invertible. We sketch the proof of these results in Appendix C. These conditions on P are probably not essential.

In a recent paper [25], Browne and DePierro introduce a new algorithm, the "row-action maximum likelihood algorithm" (RAMLA), along with a BI generalization. These algorithms are relaxed versions of the EMART and BI-EMML. Because they are primarily interested in the inconsistent case, they introduce strong underrelaxation and obtain a generalization of Theorem 2. Our approach is to go in the opposite direction, to rescale the equations with large multipliers to speed up convergence to the LC and then to use feedback.

VI. Conclusions

When the system $Ax = b$ has no solutions ART produces a limit cycle, from which a new b vector can be obtained. Applying ART to this new system leads to a second limit cycle, a next b, and so on. For each m the sequence $\{K^{-1}(z^{0,m} + z^{1,m} + \cdots + z^{K-1,m})\}$ converges to the least squares solution x^* of $Ax = b$, as K goes to infinity.

In the consistent case, in which the system of equations $y = Px$ has a nonnegative solution, BI-SMART converges to that solution minimizing the distance KL(x, x^0), where x^0 is the starting vector. If x^0 is constant, then the MART converges to the maximum Shannon entropy solution, as Lent has shown. In the inconsistent case, in simulations, subsequences of BI-SMART corresponding to complete cycles converge, but to distinct vectors, which we call the limit cycle. We prove that these vectors have a common support S with cardinality less than the number of equations. The resulting limit cycle can be used to obtain a new y. Restarting BI-SMART using this new y and restricting to the support S, we obtain a third y, and so on. This feedback results in a sequence of limit cycles that converges to the single vector x' that minimizes the distance KL(Px, y), subject to the constraint that x be supported on the (nonempty) intersection of the limit cycle support sets. A similar result holds without the support constraint.

The feedback of the new y amounts to modifying the hyperplanes. For nonlinear Bregman iteration schemes, it has been shown that this is the proper way to view relaxation. Hence, the result we obtain here is the MART analog of the theorem of Censor *et al.* [16] on the behavior of ART when the relaxation goes to zero.

Our main purpose here has been to demonstrate that the limit cycles produced by ART, BI-SMART, and BI-EMML retain sufficient information to reconstruct from them approximate solutions to the original problem. The feedback method was introduced primarily as a tool to further the theoretical development. We are currently investigating the effectiveness of the feedback approach as part of a practical image reconstruction method. Such an algorithm would probably also involve regularization for smoothness. Preliminary results have encouraged us to believe that feedback does have practical uses. Encouraging results have also been reported in connection with the OS-EM algorithm in [24].

Appendix A

Proof of Theorem 1: Consider two distinct ART sequences, $x^{k+1} = (x^k)'$, beginning at x^0 and $z^{k+1} = (z^k)'$, beginning at z^0; at each step the mth equation is used to define the successor, where $m = k(\text{mod}\, M) + 1$. We then have

$$L^2(x^0, z^0) - L^2(x^M, z^M) = (2t - t^2) \sum_m L^2(Ax_m^{m-1}, Az_m^{m-1}) \tag{A1}$$

in order for the right side of (A1) to be nonnegative we need to require that the relaxation parameter t be in [0, 2]. For t in (0, 2) the ART is a contractive mapping. For any vector x we have

$$L^2(x, z^0) - L^2(x, z^M) = (2t - t^2)\left\{\sum_m L^2(Ax_m, Az_m^{m-1}) - L^2(Ax, b)\right\} + (2t - 2t^2) \sum_m (b_m - Ax_m)(b_m - Az_m^{m-1}). \tag{A2}$$

From (A1) we obtain

$$L^2(z^M, z^0) - L^2(z^{2M}, z^M) = (2t - t^2) \sum_m L^2(Az_m^{M+m-1}, Az_m^{m-1}) \tag{A3}$$

it follows that the sequence $\{L^2[z^{(k+1)M}, z^{kM}]\}$ is decreasing and that the difference sequence converges to zero. Therefore, we have that the sequence $\{\sum_m L^2[Az_m^{(k+1)M+m-1}, Az_m^{kM+m-1}]\}$ converges to zero. The sequence $\{z^{kM}\}$ is bounded, as we shall show below, so has at least one cluster point, say z^*. We consider the collection of successors, beginning with $z^* = z^{*+0}$, so that $z^{*+k} = (z^{*+k-1})'$. Since the sequence $\{\sum_m L^2(Az_m^{(k+1)M+m-1}, Az_m^{kM+m-1})\}$ converges to zero we have that $A(z^{*+M+m-1})_m = A(z^{*+m-1})_m$ for each m.

Using (A2) to get $L^2(x, z^*) - L^2(x, z^{*+M})$ and $L^2(x, z^{*+M}) - L^2(x, z^{*+2M})$, then taking $x = z^{*+M}$, and recalling $Az_m^{*+M+m-1} = Az_m^{*+m-1}$, we find that $z^* = z^{*+M}$, so that the cluster point z^* is a fixed point of the cyclic ART iteration. Now applying (A1) with $x^0 = z^*$ we find that $\{L^2(z^*, z^{kM})\}$ is decreasing. Since a subsequence goes to zero, the whole sequence must converge to z^*. For each $m = 1, 2, \cdots, M-1$ we have $\{z^{kM+m}\}$ converging to z^{*+m}. The set of vectors $\{z^*, z^{*+1}, \cdots, z^{*+M-1}\}$ is the limit cycle (LC).

Proof of Theorem 2: The sequence $\{z^{k,0}, k = 0, 1, \cdots\}$ is bounded, so let z^* be one of its cluster points. Then by (4) z^* is a cluster point for each of the sequences $\{z^{k,m}$, m fixed, $k = 0, 1, \cdots\}$. We have

$$t_k^{-1}(z_n^{k,m} - z_n^{k,m-1}) = A_{m,n}(b_m - Az_m^{k,m-1}). \quad \text{(A4)}$$

Summing over m on both sides gives zero on the left and the nth entry of $A^T(b - d^k)$ on the right, where T denotes matrix transpose and $d_m^k = Az_m^{k,m-1}$. It follows that $A^Tb = A^Td^k$ for each k. Taking k to infinity gives $A^Tb = A^TAz^*$, since the sequence $\{d^k\}$ clusters at Az^*. But $A^Tb = A^TAz^*$ implies that z^* is the least squares solution. Since the least squares solution is unique, by assumption, the cluster point z^* is then unique and the limit cycles must converge to $z^* = x^*$.

Proof of Theorem 3: Let c be the vector with entries $c_m = Az_m^{m-1}$. Then as we saw above, $A^Tb = A^Tc$, so that the systems $Ax = b$ and $Ax = c$ have the same least squares solution. We can then write $b = Ax^* + w.c = Ax^* + v$, where $A^Tw = A^Tv = 0$, and $\|b\|^2 = \|Ax^*\|^2 + \|w\|^2$, $\|c\|^2 = \|Ax^*\|^2 + \|v\|^2$, with $\|b\|^2 = L^2(b, 0)$. A simple calculation shows that $\|z^m\|^2 - \|z^{m-1}\|^2 = (b_m)^2 - (c_m)^2$ for each m. Summing over m on both sides gives zero on the left and $\|b\|^2 - \|c\|^2$ on the right. Therefore, $\|b\|^2 = \|c\|^2$ and $\|w\|^2 = \|v\|^2$. Since $M = N + 1$ the null space of A^T has dimension one. Since both w and v are in this null space they are equal or $v = -w$. We assume no solution of $Ax = b$ exists, so $v \neq w$; therefore, $v = -w$. Now from (5), we have $L^2(x^*, z^m) - L^2(x^*, z^{m-1}) = (v_m)^2 - (w_m)^2 = 0$.

Proof of Theorem 4: Denote by P_m^k the orthogonal projection onto the hyperplane consisting of all x such that $Ax_m = b_m^k$. Then a simple calculation shows that

$$\begin{aligned} &K^{-1}(z^{0,m} + z^{1,m} + \cdots + z^{K-1,m}) \\ &\quad - K^{-1}(z^{0,m-1} + z^{1,m-1} + \cdots + z^{K-1,m-1}) \\ &= K^{-1}(P_m^0 - P_m^K)x^*. \end{aligned} \quad \text{(A5)}$$

Since the right side of (A5) goes to zero as K increases, it follows that both of the sequences $\{K^{-1}(z^{0,m} + z^{1,m} + \cdots + z^{K-1,m})\}$ and $\{K^{-1}(z^{0,m-1} + z^{1,m-1} + \cdots + z^{K-1,m-1})\}$ have the same set of cluster points. Taking the product of A with the first of these sequences, and then its mth entry, we get

$$\begin{aligned} &\{A[K^{-1}(z^{0,m} + z^{1,m} + \cdots + z^{K-1,m})]_m\} \\ &= \{K^{-1}(b_m^0 + b_m^1 + \cdots + b_m^{K-1})\}. \end{aligned} \quad \text{(A6)}$$

Now let z^* be any cluster point of the sequence $\{K^{-1}(z^{0,m} + z^{1,m} + \cdots + z^{K-1,m})\}$. Then Az_m^* is a cluster point of the right side of (A6), for each m. Since $b^k = Ax^* + w^k$ for each k, $Az^* - Ax^*$ is a cluster point for the sequence $\{K^{-1}(w^0 + w^1 + \cdots + w^{K-1})\}$. Therefore, $Az^* - Ax^*$ is in the null space of A^T and in the range of A at the same time; $Az^* - Ax^*$ must be zero, then. Since the least squares solution is unique, it follows that $z^* = x^*$ and that the z^* is in fact the limit, as claimed.

Proof of Theorem 5: Let C be the set of all nonnegative linear combinations of the columns of A. Since C is convex and Ax is the member of C closest to b in the Euclidean distance, Ax is unique, even if x may not be. For any c in C we know that the inner product $\langle b - Ax, c - Ax\rangle \leq 0$. Selecting $c = Ax + A^n$, where A^n is the nth column of A, we get $\langle b - Ax, A^n\rangle \leq 0$, for each n. Then selecting c to be $c = Ax - x_nA^n$, we get $x_n\langle b - Ax, A^n\rangle \geq 0$. It follows that $x_n\langle b - Ax, A^n\rangle = 0$ for all n. Let B be the matrix obtained from A by deleting the nth column if $x_n = 0$ for all optimal x. Then we must have $B^T(b - Ax) = 0$. If B has at least as many columns as it has rows then B^T is one-to-one and $Ax = b$. Consequently, B has fewer columns than it has rows; letting S be the set of all n corresponding to columns of A that were not deleted, the theorem follows.

Proof that $\{z^{kM}\}$ Is Bounded: For $k = 0, 1, \cdots$ and $m = 1, \cdots, M$ set $v^{k,m} = (Az^{kM+m-1})_m$ and let v^k be the vector with entries $v^{k,m}$. There are two cases to consider: i) the sequence of vectors $\{v^k\}$ has an infinite bounded subsequence; or ii) it does not. We treat case ii) first. Using (A2) with $x = 0$ we see that $\|z^{kM}\|^2 - \|z^{(k+1)M}\|^2$ is eventually decreasing, so $\{z^{kM}\}$ is bounded. Case i) is more complicated. To avoid cumbersome notation, we shall assume that the entire sequence $\{v^k\}$ is bounded, although we shall not use the fact that it is the entire sequence in what follows. For each $m = 1, \cdots, M$ and $k = 0, 1, \cdots$, let $K^m = \{x|Ax_m = 0\}$, $H^{k,m} = \{x|Ax_m = A(z^{kM+m})_m\}$ and $p^{k,m}$ the orthogonal projection of 0 onto the hyperplane $H^{k,m}$. We then have $H^{k,m} = K^m + p^{k,m}$. For each k and m the distance $L^2(z^{kM+m-1}, z^{kM+m}) = L^2(b_m, v^{k,m})$, so $\{L^2(z^{kM+m-1}, z^{kM+m})\}$ is bounded for each fixed m. The projections $\{p^{k,m}\}$ are also bounded for each fixed m. If $\{z^{kM+m-1}\}$ and $\{z^{kM+m}\}$ are unbounded then the sequences $\{x^{kM+m-1} = z^{kM+m-1} - p^{k,m-1}\}$ and $\{x^{kM+m} = z^{kM+m} - p^{k,m}\}$ are unbounded in K^{m-1} and K^m, respectively, with $\{L^2(x^{kM+m-1}, x^{kM+m})\}$ bounded. We let $u^{k,m}$ and $u^{k,m-1}$ be the unit vectors $x^{kM+m}/\|x^{kM+m}\|$ and $x^{kM+m-1}/\|x^{kM+m-1}\|$, respectively. Then we have $\|x^{kM+m-1} - x^{kM+m}\|^2 \geq \|x^{kM+m-1} - \|x^{kM+m-1}\|\langle u^{k,m-1}, u^{k,m}\rangle u^{k,m}\|^2$, since $\|x^{kM+m-1}\|\langle u^{k,m-1}, u^{k,m}\rangle u^{k,m}$ is the projection of x^{kM+m-1} onto the span of $u^{k,m}$. Therefore, $\{\|x^{kM+m-1}\|\,\|u^{k,m-1} - \langle u^{k,m-1}, u^{k,m}\rangle u^{k,m}\|\}$ is bounded.

This means that $\{1 - \langle u^{k,m-1}, u^{k,m}\rangle\}$ goes to zero and that $\{u^{k,m}\}$ and $\{u^{k,m-1}\}$ have the same cluster points. From this, we conclude that K^m and K^{m-1} have nonzero intersection. Since m is arbitrary, it follows that the null space of matrix A must have nonzero members. Since A has full rank and $M > N$, this cannot happen.

Appendix B

Proof of Lemma 1: Fix j and assume that it has been shown that $z_j^k \leq C_j$; we show then that $z_j^{k+1} \leq C_j$. Let $n = k(\mathrm{mod}\ N) + 1$. We consider only those i in S_n for which $P_{i,j} > 0$. Then, $\log(z_j^{k+1}) \leq \log(z_j^k) + \sum^n P_{i,j} \log(y_i/P_{i,j}z_j^k)$, so that $\log(z_j^{k+1}) \leq [1 - \sum^n P_{i,j}] \log(z_j^k) + \sum^n P_{i,j} \log(y_i/P_{i,j})$. It follows that $z_j^{k+1} \leq \max\{z_j^k, y_i/P_{i,j}, i \text{ in } S_n\}$.

Proof of Theorem 6: We have $\sum_i P_{i,j} \log(y_i/w_i) = \lim \log[z_j^{(m+1)N}/z_j^{mN}] = 0$ for each j in S. Let Q be obtained from P be deleting the jth column of P whenever j is not in S, the support of the limit cycle vectors. Then we can write $Q^T u = 0$, where u is the vector with entries $u_i = \log(y_i/w_i)$. But since Q^T has full rank, it must have fewer rows than columns, otherwise $u = 0$. If $u = 0$ then $y_i = w_i$ for all i, from which it follows that the LC consists of a single vector, which must then solve $y = Px$.

Proof of Lemma 2: If not, then $0 < \sum_i P_{i,j} \log(y_i/w_i) \leq +\infty$, for some j. Then there is M and $e > 0$ such that, for $m \geq M$, we have $\sum_i P_{i,j} \log(y_i/Pz_i^{mN+n-1}) \geq e$, where n and i are related by i in S_n. For $\{x^m = z^{mN}\}$ we have $x_j^m < \exp(e)x_j^m \leq x_j^{n+1} < \exp(e)x_j^{m+1}, \cdots$, violating the condition that $\{x^m\}$ is bounded. So $\sum_i P_{i,j} \log(y_i/w_i) \leq 0$ for all j. For j in S, the equality follows from the proof of Theorem 6.

Proof of Theorem 7: It follows from Lemma 3 that we have $\sum_n G_n(z^{*,n}, z^{*,n-1}) = 0$. With our assumptions on P, this means that $z^{*,n} = z^{*,n+1}$, for all n. Then we can conclude that $z^{*,n} = z^*$, for all n, and so $Pz_i^{*,n-1} = Pz_i^*$ for all i. Since $\{\mathrm{KL}(Px, y^k)\}$ is decreasing for all x, it follows that $\{y^k\}$ converges to Pz^*. If $J < I$ and if each limit cycle has full support then z^* is the unique nonnegative minimizer of the quantity $\mathrm{KL}(Px, y)$, so it is the limit of the SMART iteration: This is true because we have equality in (13) at each step, so minimizing $\mathrm{KL}(Px, y)$ is equivalent to minimizing $\mathrm{KL}(Px, Pz^*)$.

Appendix C

Once again, we assume the existence of the LC(k). With $y_i^{k+1} = Pz_i^{k,n-1}$, where the n is chosen so that i is in the nth subset, we have the following.

1) $\sum_i y_i^{k+1} = \sum_i y_i^k$: the difference is $\sum_n\{\sum_j z^{k,n} - \sum_j z^{k,n-1}\} = 0$.
2) From 1) we have $\{y^k\}$ is bounded.
3) There is constant B such that, for all k, m, n, and j, $z_j^{k,mN+n} < B$. For the proof, let $B_i \geq y_i^k$ for all k. Then let $M_j = \max\{B_i/P_{i,j}\}$ over all i for which $P_{i,j} > 0$. Then let $C_j = \max\{M_j, z_j^0\}$ and $B = \max\{C_j\}$.
4) For all j and k we have $\prod_n\{(1 - \sum^n P_{i,j}) + \sum^n P_{i,j}y_i^k/y_i^{k+1}\} \leq 1$.

   Proof: If not, then for some j and k the product is $\geq 1+a$, for some $a > 0$. Then, there is M such that, for all $m \geq M$, $\prod_n\{(1 - \sum^n P_{i,j}) + \sum^n P_{i,j}y_i^k/Pz_i^{k,mN+n-1}\} \geq 1 + a/2$. But this product is $z_j^{k,(m+1)N}/z_j^{k,mN}$, and $z_j^{k,mN}$ converges to $z^{k,0}$. Note that the sequence $z^{k,mN+n-1}$ denotes the BI-EMML iterations that lead to the LC(k), not to the members of LC(k) itself.
5) It follows from 4) that for all $x \geq 0\{\mathrm{KL}(Px, y^k)\}$ is decreasing and that the sequence $\{P^T(\log y^k)\}$ is convergent.
6) Therefore, if P^T is one-to-one, $\{y^k\}$ is convergent.
7) $\sum_n \mathrm{KL}(z^{k,n-1}, z^{k,n}) \leq \mathrm{KL}(y^{k+1}, y^k)$, so if $\{y^k\}$ converges then every cluster point of the sequence $\{z^{k,0}\}$ generates a cycle that is a singleton, say z^*, with $y^\infty = Pz^*$. If this equation determines z^* uniquely then the LC(k) must converge to z^*.

Acknowledgment

The author thanks Profs. J. Graham-Eagle of UML and Y. Censor of the University of Haifa for helpful discussions on these matters.

References

[1] C. Byrne, "Block-iterative methods for image reconstruction from projections," *IEEE Trans. Image Processing,* vol. 5, pp. 792–794, May 1996.
[2] G. Herman and L. Meyer, "Algebraic reconstruction techniques can be made computationally efficient," *IEEE Trans. Med. Imag.,* vol. 12, pp. 600–609, Sept. 1993.
[3] H. Guan and R. Gordon, "A projection access order for speedy convergence of ART (algebraic reconstruction technique): A multilevel scheme for computed tomography," *Phys. Med. Biol.,* vol. 39, pp. 2005–2022, 1994.
[4] 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, pp. 471–481, 1970.
[5] K. Tanabe, "Projection method for solving a singular system of linear equations and its applications," *Numer. Math.,* vol. 17, pp. 203–214, 1971.
[6] M. Trummer, "SMART—An algorithm for reconstructing pictures from projections," *J. Appl. Math. Phys.,* vol. 34, pp. 746–753, Sept. 1983.
[7] C. Byrne, "Iterative image reconstruction algorithms based on cross-entropy minimization," *IEEE Trans. Image Processing,* vol. 2, pp. 96–103, Jan. 1993.
[8] ——, "Erratum and addendum to 'Iterative image reconstruction algorithms based on cross-entropy minimization,'" *IEEE Trans. Image Processing,* vol. 4, pp. 225–226, Feb. 1995.
[9] ——, "Iterative reconstruction algorithms based on cross-entropy minimization," in *Image Models (and Their Speech Model Cousins),* vol. 80, S. Levinson and L. Shepp, Eds. New York: Springer-Verlag, 1996.
[10] ——, "Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative (RBI) methods," *IEEE Trans. Image Processing,* to be published.
[11] S. Kaczmarz, "Angenaeherte Aufloesung von Systemen linearen Gleichungen," *Bull. Acad. Polon. Sci. Lett.,* vol. A35, pp. 355–357, 1937.
[12] G. Herman, A. Lent, and P. Lutz, "Relaxation methods for image reconstruction," *Commun. Assoc. Comput. Mach.,* vol. 21, pp. 152–158, 1978.
[13] M. Trummer, "Reconstructing pictures from projections: On the convergence of the ART algorithm with relaxation," *Computing,* vol. 26, pp. 189–195, 1981.
[14] P. Eggermont, G. Herman, and A. Lent, "Iterative algorithms for large partitioned linear systems, with applications to image reconstruction," *Linear Algebra Appl.,* vol. 40, pp. 37–67, 1981.

[15] A. Dax, "The convergence of linear stationary iterative processes for solving singular unstructured systems of linear equations," *SIAM Rev.,* vol. 32, pp. 611–635, Dec. 1990.

[16] Y. Censor, P. Eggermont, and D. Gordon, "Strong underrelaxation in Kaczmarz's method for inconsistent systems," *Numer. Math.,* vol. 41, pp. 83–92, 1983.

[17] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Stat.,* vol. 22, pp. 79–86, 1951.

[18] A. Lent, "A convergent algorithm for maximum entropy image restoration with a medical x-ray application," in *Image Analysis and Evaluation,* R. Shaw, Ed. Washington, DC: SPSE, 1977, pp. 238–243.

[19] A. R. DePierro and A. N. Iusem, "A relaxed version of Bregman's method for convex programming," *J. Optimiz. Theory Appl.,* vol. 51, pp. 421–440, 1986.

[20] Y. Censor, "Finite series-expansion reconstruction methods," *Proc. IEEE,* vol. 71, pp. 409–419, Mar. 1983.

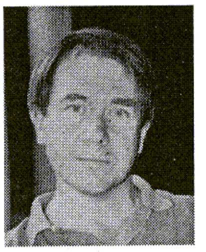
[21] G. Herman, "Applications of maximum entropy and Bayesian optimization methods to image reconstruction from projections," in C. P. Smith and W. T. Grandy, Jr., Eds., *Maximum Entropy and Bayesian Methods in Inverse Problems.* D. Reidel, 1985., pp. 319–338.

[22] C. P. Smith and W. T. Grandy, Jr., Eds., *Maximum Entropy and Bayesian Methods in Inverse Problems.* D. Reidel, 1985.

[23] H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.* vol. 13, pp. 601–609, Dec. 1994.

[24] W. Xuan, "Evaluating a feedback technique in image reconstruction from projections," Tech. Rep., Dept. Statist., MacQuarie Univ., Sydney, Australia, Dec. 1993.

[25] J. Browne and A. DePierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Trans. Med. Imag.,* vol. 15, pp. 687–699, Oct. 1996.



Charles L. Byrne (M'87) received the B.S. degree from Georgetown University, Washington, DC, in 1968 and the M.A. and Ph.D. degrees from the University of Pittsburgh, Pittsburgh, PA, in 1970 and 1972, respectively, all in mathematics.

From 1972 to 1986, he was a member of the Department of Mathematics, Catholic University of America, Washington, DC, serving as Department Chairman from 1983 to 1986. Since 1986, he has been with the Department of Mathematical Sciences, University of Massachusetts, Lowell, serving as Department Chairman from 1987 to 1990. He has served as consultant to the U.S. Naval Research Laboratory and to the Australian Department of Defense in the area of underwater acoustic signal processing. Since 1990, he also served as consultant in tomographic image reconstruction to the Department of Nuclear Medicine, University of Massachusetts Medical Center, Worcester.