Iterative Image Reconstruction Algorithms Based on Cross-Entropy Minimization

Charles L. Byrne, *Member, IEEE*

***Abstract*— The cross-entropy (or Kullback–Leibler) distance between two nonnegative vectors a and b is $\mathrm{KL}(a,b) = \sum a_n \log(a_n/b_n) + b_n - a_n$. Several well-known iterative algorithms for reconstructing tomographic images lead to solutions that minimize certain combinations of KL distances, and can be derived from alternating minimization of related KL distances between convex sets; these include the expectation maximization (EM) algorithm for likelihood maximization (ML), and the Bayesian maximum *a posteriori* (MAP) method with gamma-distributed priors, as well as the multiplicative algebraic reconstruction technique (MART). Each of these algorithms can be viewed as providing approximate nonnegative solutions to a (possibly inconsistent) linear system of equations, $y = Px$.**

In almost all cases, the ML problem has a unique solution (and so the EM iteration has a limit that is independent of the starting point) unless the system of equations $y = Px$ has a nonnegative solution, regardless of the dimensions of y and x.

We introduce the "simultaneous" MART (SMART) algorithm and prove convergence: for $0 < \alpha < 1$, SMART converges to the $x \geq 0$ for which $\alpha\mathrm{KL}(Px, y) + (1-\alpha)\mathrm{KL}(x, p)$ is minimized, where p denotes a prior estimate of the desired x; for $\alpha = 1$, the SMART algorithm converges in the consistent case (as does MART) to the unique solution of $y = Px$ minimizing $\mathrm{KL}(x, x^0)$, where x^0 is the starting point for the iteration, and in the inconsistent case, to the unique nonnegative minimizer of $\mathrm{KL}(Px, y)$.

I. Introduction and Background

In [1], Rockmore and Macovski suggest that improvements in imaging from emission tomography (ET) data can be achieved by making greater use of the underlying Poisson model for the emissions in the design of the reconstruction algorithm; specifically, they suggest that the desired vector of intensities (mean values of the Poisson random variables) be obtained through maximum likelihood (ML) parameter estimation. In subsequent papers [2]–[6], a number of authors have considered the use of the "expectation maximization" (EM) algorithm to iteratively calculate the ML solution. More recently, Bayesian maximum *a posteriori* (MAP) reconstruction has been offered as a way to increase the smoothness of the reconstructions and to permit the inclusion of prior information in the reconstruction process [7]–[15].

Manuscript received November 19, 1991; revised May 26, 1992. This work was supported by the National Cancer Institute under Grants CA42165 and CA51071. Its contents are solely the responsibility of the author and do not necessarily represent the official views of the National Cancer Institute. The associate editor coordinating the review of this paper and approving it for publication was Dr. Yoram Bresler.

The author is with the Department of Mathematics, University of Massachusetts at Lowell, Lowell, MA 01854.

IEEE Log Number 9205207.

The cross-entropy (or Kullback–Leibler) distance [16] between two nonnegative vectors a and b is $\mathrm{KL}(a,b) = \sum a_n \log(a_n/b_n) + b_n - a_n$. With the support (b) defined to be the set of indexes n for which $b_n > 0$, we have $\mathrm{KL}(a,b) = +\infty$ unless support(a) is contained in support(b). As noted by Titterington [17], maximizing the likelihood function derived from the Poisson model is equivalent to minimizing $\mathrm{KL}(y, Px)$, the Kullback–Leibler distance between the $I \times 1$ vector y of counts at each detector (I is the number of detectors) and the expected number Px, where x is the nonnegative $J \times 1$ vector whose jth entry x_j denotes the mean value of the Poisson emissions at pixel j, and the $I \times J$ matrix P has entries $P_{i,j}$ = probability that a particle emitted at pixel j will be detected at detector i. Minimizing the distance $\mathrm{KL}(y, Px)$ over x in the nonnegative orthant can be achieved through an iterative scheme involving alternating projections onto certain convex sets [18]; this is a useful framework within which to understand and extend the EM/ML algorithm.

In the ET case, the count data y are contaminated by Poisson noise, which is the difference between the actual counts y and the expected counts Px. As the dosage increases, and with it the particle counts, the signal-to-noise ratio (SNR) improves, but for realistic dosages and count levels, noise is significant. To avoid overfitting to noisy data, as well as to increase the smoothness of the reconstructed image, several authors have turned to Bayesian maximum *a posteriori* (MAP) methods. These can be viewed as regularization procedures, and generally lead to the minimization of functionals of the form $F(x) = \alpha\mathrm{KL}(y, Px) + (1-\alpha)D(x)$, where $D(x)$ is some "penalty" function that forces upon x a degree of smoothness or nearness to a prior estimate, and $0 < \alpha < 1$. If the $D(x)$ is carefully chosen, the alternating projections approach can be extended to this case. In particular, if we use $D(x) = \mathrm{KL}(p, x)$, where p is a nonnegative prior estimate of the vector x, then the iterative method of Lange, *et al.* [9] can be obtained through alternating minimization.

In the ET applications, J = the number of pixels typically is roughly equal to I = the number of detectors—although J is sometimes increased to make the resolution appear better—and noise is spread, more or less, over all of the data values in y; we seek a nonnegative solution of an essentially overdetermined and noisy linear system of equations. In other applications, such as those involving power spectrum estimation from short time series or array processing, the situation is different. Here, resolution can be the main problem; that is, J is typically greater than I, and the data are limited in amount, not in quality. For example, in acoustic array processing, spatial resolution is limited by the aperture or the greatest separation between sensors (in units of wavelength), as well as by the higher statistical variability of those estimated cross-sensor correlations corresponding to greater spatial separation, for which relatively few representatives are available. In such cases, the problem is to find a suitable nonnegative solution of an underdetermined (and mildly noisy) linear system. One approach that has been considered for such problems is the maximization of entropy or related functionals, subject to (hard or soft) data consistency constraints [19]–[31], [38], [39].

Having chosen a prior estimate p of x, we can minimize the functionals $\mathrm{KL}(p, x)$ or $\mathrm{KL}(x, p)$, subject to nonnegativity and data consistency ($Px = y$). If p is a constant vector, the first approach is equivalent to the maximization of Burg entropy, while the second is equivalent to the maximization of Shannon entropy. If one wishes to relax the data consistency constraints, one can instead minimize a functional of the form $\alpha D(Px, y) + (1-\alpha)\mathrm{KL}(p, x)$ or $\alpha D(Px, y) + (1-\alpha)\mathrm{KL}(x, p)$, where the relaxation term $D(Px, y)$ denotes some measure of distance from Px to y, thereby forcing some approximate data consistency on the solution. If the relaxation term is carefully chosen, these functionals can also be minimized via alternating projections. If, in the first case above, we select $D(Px, y) = \mathrm{KL}(y, Px)$, we have the same functional considered earlier. If we select $D(Px, y) = \mathrm{KL}(Px, y)$ in the second case, we obtain a "simultaneously updated" version of the MART algorithm considered by Gordon *et al.*, Censor, Lent, Herman, and others [32]–[35], [39], which we call SMART.

The multiplicative algebraic reconstruction technique (MART) introduced into image processing by Gordon *et al.* [32] and discussed by Herman, Lent, and Censor, among others [33]–[35], [39], is another method that can best be understood and extended within the alternating projections framework. Lent has shown that MART converges to the solution maximizing Shannon entropy $\mathrm{SE}(x) = -\Sigma x_j \log x_j$ in the consistent case, that is, whenever the system of equations $y = Px$ has a nonnegative solution. In [39], Herman notes that in the inconsistent case, MART may not maximize entropy. In [33], Censor remarks that the behavior of MART in the inconsistent case is not known. We show later that MART can be derived as an alternating projections algorithm. In the consistent case, the MART iterative scheme converges to the unique nonnegative solution of $y = Px$ minimizing $\mathrm{KL}(x, x^0)$, where x^0 denotes the starting point for the iteration; if the x^0 is chosen to be a constant vector, then the iterates converge to the maximum Shannon entropy solution, as Lent has shown. In the inconsistent case, the behavior of MART is more complicated than that of SMART; some partial results are given at the end of Section VII, with details to be presented in [41]. MART has been generalized to block-iterative methods [43]–[46], and unrelaxed SMART is nearly a special case as well; convergence proofs for the inconsistent case have not been obtained.

In [6], it is claimed that there are multiple solutions minimizing $\mathrm{KL}(y, Px)$ whenever $J > I$; this is not true. In as yet unpublished work [42], Shepp and Vanderbei conclude, on the basis of simulation studies and despite the claim in [6], that in the inconsistent case, the ML solution should be expected to be "essentially unique, in low count cases." As we shall show, essentially the only time multiple solutions are possible is in the consistent case, in which there is a nonnegative solution of the equations $y = Px$; in the inconsistent case, the EM iterative method converges to the unique ML solution, regardless of the starting point, and this ML solution has at most I-1 nonzero entries. For the case of low SNR and $J > I$, the ML solution has the choice of fitting the noisy data exactly (consistent case) or having no more than I-1 nonzero entries. From this, we can see why the ML solution tends to be nonsmooth.

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 are necessary for $x \geq 0$ to be a global minimizer of $f(x)$:

$$\partial f(x)/\partial x_j = 0, \qquad \text{if } x_j > 0 \tag{1}$$

$$\partial f(x)/\partial x_j \geq 0, \qquad \text{if } x_j = 0. \tag{2}$$

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, (1) and (2) are also sufficient for x to be a global minimizer.

II. The Algebraic Reconstruction Problem

We begin by deriving the algebraic problem from the likelihood maximization based on the Poisson model of emitters. In emission tomography, one imagines J spatial locations (pixels or voxels) such that the emission counts per unit of time at the various locations are independent and at the jth location constitute a Poisson random variable with mean value x_j. The counts obtained at detector i, denoted y_i, are also independent across detectors and Poisson, with mean values $E(y_i) = Px_i$, where $Px_i = \sum P_{i,j}x_j$, $P_{i,j}$ is the probability that a particle emitted at location j will be detected at sensor i, and the sum is over the repeated index, a convention we adopt throughout the paper. We shall assume, for convenience, that for each j, $\sum P_{i,j} = 1$, where the sum is over i; in ET, not all emitted particles are detected, so some rescaling of the original probabilities and redefinition of what is meant by x is required to achieve this simplification. Given the data vector y, the log-likelihood function to be maximized with respect to x is

$$\mathrm{LL}(x) = \sum[-Px_i + y_i \log Px_i - \log(y_i!)]. \tag{3}$$

The Kuhn–Tucker necessary and sufficient conditions for a maximum of $\mathrm{LL}(x)$ are then

$$1 = \sum P_{i,j}y_i/Px_i, \qquad \text{if } x_j > 0 \tag{4}$$

$$1 \geq \sum P_{i,j}y_i/Px_i, \qquad \text{if } x_j = 0. \tag{5}$$

Consequently, if x is an ML solution, then we have, for all j,

$$x_j = x_j \sum P_{i,j}y_i/Px_i; \tag{6}$$

it follows that $\sum x_j = \sum y_i = y_+$. It is easy to see that x is an ML solution if and only if x also minimizes $\mathrm{KL}(y, Px)$

over nonnegative x; the ML approach seeks to minimize the distance between y and Px, constrained only by the nonnegativity of the x.

We shall assume throughout that the matrix P has been constructed so that P, as well as any submatrix Q obtained from P by deleting columns, has full rank. In particular, if Q is $I \times K$ and $K \geq I$, then Q has rank I. The columns of P are vectors in the nonnegative orthant of I-dimensional space. When attenuation, detector response, and scattering are omitted from the design of the projection matrix P, it can sometimes happen that P or some submatrices Q can fail to be full rank. However, the slightest perturbation of the entries of such a P will almost surely produce a new P having the desired full-rank properties. What happens in the rank-deficient cases is isolated pathology; some condition on rank is clearly needed for Proposition 1 below to hold, as the following example illustrates. Let $I = J = 2$, let P have rows $(1-\varepsilon, 1-\varepsilon)$, $(\varepsilon, \varepsilon)$ for some ε with $1/2 > \varepsilon > 0$, and let $y = (1,1)^T$. The ML solutions are not unique: any $0 \leq x = (x_1, x_2)^T$ such that $x_1 + x_2 = 2$ will maximize likelihood; then $Px = (2(1-\varepsilon), 2\varepsilon)^T$ is the projection vector. The rank of the matrix P is one, not two, so P is rank-deficient and Proposition 1 fails.

Let $C = \{\text{all } Px | x \geq 0\}$ be the convex cone built from the columns of P, H the convex hull of the columns of P, and $A = \{z = (z_i) | \sum z_i = 1\}$. It is well known [40, p. 20] that each member of H can be written as a convex combination of $I+1$ or fewer columns of P; since the column sums in P are all one, it follows that H is contained in the I-1-dimensional affine subspace A, so that any member of H can be written as a convex combination of I or fewer columns of P. Our assumption says that no column of P is a linear combination of I-1 or fewer other columns of P, so therefore no column of P lies in the set of measure zero of convex combinations of I-1 or fewer other columns of P.

Given the data vector y, also in the nonnegative orthant of I-dimensional space, it may or may not be in C. As we increase the number of pixels J and add columns to the matrix P, the cone C expands and it becomes easier for y to be in C.

Note that the vector Px^{ML} minimizing $\mathrm{KL}(y, Px)$ over nonnegative x is always unique, even if the x^{ML} is not. This follows from the strict convexity of $\mathrm{KL}(y, w)$ as a function of w: if $\mathrm{KL}(y, Px) = \mathrm{KL}(y, Pz)$ for some $Px \neq Pz$, then for any t in (0,1), we have $\mathrm{KL}(y, tPx + (1-t)Pz) = \mathrm{KL}(y, P(tx + (1-t)z)) < \mathrm{KL}(y, Px)$.

Proposition 1: Let $J \geq I$, and let $M = \{\text{all for } x \geq 0 \text{ for which } \mathrm{KL}(y, Px) \text{ attains its minimum value}\}$. Then either

1) the minimum of $\mathrm{KL}(y, Px)$ is zero and $y = Px$ has nonnegative solutions, or

2) there is a subset of S of $\{1, 2, \cdots, J\}$ having I-1 or fewer members such that if x is in M, then $x_j > 0$ only if j is in S.

In the former case, there may be multiple solutions; M may contain more than one vector. In the latter case, $M = \{x^{\mathrm{ML}}\}$, x^{ML} the unique ML solution.

Proof: Let Q be the $I \times K$ matrix ($K \leq J$) obtained from P by deleting the jth column of P whenever $x_j = 0$ for all x in M. We know that Px is the same for all x in M; let u be the vector $u = (u_i)$, with $u_i = y_i / Px_i$. The first Kuhn–Tucker condition (4) then says that $Q^T u = 1 = (1, \cdots, 1)^T$, with the superscript T denoting matrix transpose. Our assumption that $\sum P_{i,j} = 1$ for all j tells us that $Q^T 1 = 1$. If $K \geq I$, then, since Q has rank I, the transformation Q^T is one-to-one, from which it follows that $u = 1$, and so $y = Px$. Therefore, if there is no nonnegative solution of $y = Px$, then $K \leq I - 1$, and there is S as above. The vector Px uniquely determines x when the additional constraint is imposed that $x_j = 0$ unless j is in S.

It is known that the EM iteration always converges to an ML solution. From Proposition 1, we know that in almost all cases, the limit is independent of the starting vector unless we are in the consistent case. When $y = Px$ has nonnegative solutions, the EM iteration converges to a solution; properties of that solution, in particular, how it depends on the starting vector, will be considered later.

It has been noted by researchers in the field that the ML solution tends to be rough. It has been suggested that this is because the ML approach is a parameter estimation method, designed to estimate a few parameters from a larger data set; it breaks down when the number of parameters is as large as, or larger than, the data set. From Proposition 1, we see that adding more pixels to approximate a continuous object will not produce smoother ML reconstructions; when $J \geq I$ and the SNR is low, the ML solution must either be consistent with the noisy data y or have $J - I + 1$ zero entries. In the inconsistent case, the ML solution will select a subset of I-1 or fewer parameters to which it assigns nonzero values.

We have seen that by adopting the ML formalism, we are led to a purely algebraic problem: reconstruct $x \geq 0$ from noisy values of Px. The Poisson statistical model for the emission tomography case leads to the minimization of $\mathrm{KL}(y, Px)$, but other measures of the distance from y to Px are certainly permissible. There are many other reconstruction situations that lead to similar algebraic problems: imaging from Fourier transform data in optics or array processing, for example, or reconstructing probability densities from moment estimates. The problem we consider here is somewhat special in that we require that P be a matrix with nonnegative entries, but in many situations, the data can be transformed to achieve this. In the case of Fourier cosine or sine transform (FT) data, for example, if one of the data values is the so-called "dc" term ($y_0 = \sum x_j$), then we average y_0 and each of the other real FT values to produce a new data set, with the corresponding matrix nonnegative.

III. Regularizing the ML Solution

Because the ML solution tends to be rough, it is common to regularize to achieve a higher degree of smoothness. Instead of minimizing $\mathrm{KL}(y, Px)$, we can minimize $F(x) = \alpha \mathrm{KL}(y, Px) + (1-\alpha) D(x)$, where $0 < \alpha < 1$ and $D(x)$ is some nonnegative penalty term. Because our goal is to extend the EM/ML iterative scheme using alternating minimization, we consider a particularly convenient choice of the regularizing term: $D(x) = \mathrm{KL}(p, x)$, where $p \geq 0$ is a prior estimate of x, with $\sum p_j = \sum y_i$. Other choices for $D(x)$

have been considered in the literature [7]–[15]; this particular choice leads to an iterative algorithm previously obtained by Lange *et al.* [9] as a Bayesian MAP method.

We therefore regularize the ML solution by seeking a minimum of the functional

$$F(x) = \alpha \mathrm{KL}(y, Px) + (1-\alpha)\mathrm{KL}(p, x). \tag{7}$$

If p is a constant vector, then the second term can be replaced by the negative of Burg entropy, $\mathrm{BE}(x) = \sum \log x_j$.

In contrast to the ET case, in many applications, the problem is essentially underdetermined and noise is not the major consideration. In such cases, multiple solutions of $y = Px$ are the rule, and one must devise a methodology for selecting one from the many possibilities. Entropy maximization is one such methodology. The two entropies most frequently considered are the so-called "Burg entropy" and "Shannon entropy"; the Shannon entropy is $\mathrm{SE}(x) = -\sum x_j \log x_j$. More generally, if some prior estimate p of the x is available, one might minimize the cross entropy to this prior estimate; that is, one might minimize $\mathrm{KL}(p,x)$ or $\mathrm{KL}(x,p)$, subject to the data consistency requirement $y = Px$.

In some cases, noise in the data y prompts a relaxation of the data consistency constraint, and one minimizes a functional of the form $(1-\alpha)\mathrm{KL}(p,x) + \alpha D(y, Px)$ or of the form $(1-\alpha)\mathrm{KL}(x,p) + \alpha D(y,Px)$, where $D(y,Px)$ measures the distance from y to Px. Some may feel that almost any reasonable choice of $D(y,Px)$ will suffice to avoid the sensitivity to noise that accompanies ill-conditioned inverse problems. Here, however, we are interested in deriving iterative methods based on alternating projections, and the choice of $D(y,Px)$ is important.

We shall consider two problems: for $x \geq 0$ and $0 < \alpha \leq 1$.

Problem A: Minimize $F(x) = \alpha\ \mathrm{KL}(y, Px) + (1-\alpha)\ \mathrm{KL}(p,x)$.

Problem B: Minimize $G(x) = \alpha\ \mathrm{KL}(Px, y) + (1-\alpha)\ \mathrm{KL}(x,p)$.

In considering these particular mixtures of KL distances, we unify approaches commonly taken in the underdetermined and overdetermined cases, and we develop iterative solution methods within a single framework of alternating projections.

IV. Iterative Solutions from Alternating Projections

In what follows, we shall assume that the data vector y and the matrix P are known and fixed. We begin by defining two convex sets of IJ-dimensional vectors (see also Vardi *et al.* [6] and Csiszar and Tusnady [18]):

$$R = \{r = \{r_{i,j}\} \geq 0 | \text{ for all } i \text{ we have } \sum r_{i,j} = y_i, \text{ sum over } j\};$$

$$Q = \{q = q(x) = \{q_{i,j} = P_{i,j}x_j\}, \qquad \text{for some } x \geq 0\}.$$

Also define, for each $x \geq 0$, $r(x) = \{P_{i,j}x_j y_i / Px_i\}$ in R. For r and q in R and Q, respectively, we define $\mathrm{KL}(r,q)$ and $\mathrm{KL}(q,r)$ in the obvious manner. We also define $F(r,q)$ and $G(q,r)$ as follows:

$$F(r,q) = F(r, q(x)) = \alpha\ \mathrm{KL}(r,q) + (1-\alpha)\mathrm{KL}(p,x);$$

$$G(q,r) = G(q(x), r) = \alpha\ \mathrm{KL}(q,r) + (1-\alpha)\mathrm{KL}(x,p).$$

The alternating projections are defined as follows: let $d(r,q)$ denote either $F(r,q)$ or $G(q,r)$, and let x^0 be the starting point. Then, having obtained the kth iterate x^k and setting $q^k = q(x^k)$, perform the following.

Step 1: Minimize $d(r, q^k)$ with respect to r to get r^{k+1}.

Step 2: Minimize $d(r^{k+1}, q(x))$ with respect to $x \geq 0$ to get x^{k+1}.

We apply this iterative scheme to the two problems under consideration.

V. The Iterative Method for Problem A

Performing Step 1, with $d(r,q) = F(r,q)$, we obtain, for each j, $\log(r_{i,j}) = \log(q(x^k)_{i,j}) + c_i$ for some constants c_i. Since $\sum r_{i,j} = y_i$ for each i, we must have $r_{i,j} = P_{i,j}x_j^k y_i / Px_i^k$, so that $r^{k+1} = r(x^k)$. Now, performing Step 2, we obtain, for all j such that $x_j^{k+1} > 0$,

$$0 = \alpha\Big[\sum_i [P_{i,j} - r_{i,j}^{k+1}/x_j^{k+1}]\Big] + (1-\alpha)\big(1 - p_j/x_j^{k+1}\big). \tag{8}$$

Solving for x_j^{k+1}, we obtain

$$x_j^{k+1} = \alpha \sum_i r_{i,j}^{k+1} + (1-\alpha)p_j. \tag{9}$$

For $\alpha = 1$, we obtain the well-known EM/ML iteration [6]. For $\alpha < 1$, the iteration is that obtained by Lange *et al.* [9], using a Bayesian approach with a prior chi-squared distribution for each x_j.

Proof of convergence for the iterative methods will depend heavily on orthogonality conditions that can be viewed as versions of the Pythagorean theorem applied to the KL distance (which should be thought of as distance squared, actually). We now obtain some of these orthogonality conditions for the iteration just presented.

Lemma 1: $F(r^{k+1}, q(x)) - F\big(r^{k+1}, q(x^{k+1})\big) = \mathrm{KL}(x^{k+1}, x)$ for all $x \geq 0$.

Proof: It follows, with a little calculation, from (8).

Lemma 2: $F\big(r, q^k\big) - F\big(r^{k+1}, q^k\big) = \alpha\ \mathrm{KL}\big(r, r^{k+1}\big)$ for all r in R.

Proof: It follows immediately from the definitions and simple calculation.

Lemma 3: $\mathrm{KL}\big(x^{k+1}, x^k\big)$ goes to zero as k goes to infinity.

Proof: It follows from the two previous lemmas and the inequalities $F\big(r^k, q^k\big) \geq F\big(r^{k+1}, q^k\big) \geq F\big(r^{k+1}, q^{k+1}\big)$; since $\{F(r^k, q^k)\}$ is a decreasing sequence of nonnegative quantities, the difference sequence $\{F(r^k, q^k) - F(r^{k+1}, q^{k+1})\}$ goes to zero.

For completeness, we sketch the proof of convergence of the sequence $\{x^k\}$, along the lines of [6]. Let $\alpha \leq 1$ for now. First, because $\sum x_j^k = \sum y_i$ for each $k = 1, 2, \cdots$, it follows that the sequence is contained within a bounded subset of I-dimensional space. Then let x^* be any subsequential limit point of the sequence, so that there is a subsequence $\{x^{kn}\}$ converging to x^*. Define the vector x' by $x'_j = \alpha x_j \sum P_{i,j} y_i / Px_i + (1-\alpha)p_j$ for each j, and $x \geq 0$, with the sum over i. Say that x is a fixed point of the iteration if $x' = x$. Then the sequence $\{x^{kn+1}\}$ converges to $(x^*)'$. But then Lemma 3 gives that $(x^*)' = x^*$, so that x^* is a fixed point of the iteration, and thus satisfies the first of the two Kuhn–Tucker conditions (2) for $f(x) = F(x)$.

For the case $\alpha < 1$, we have $x_j^{k+1} \geq (1-\alpha)p_j > 0$ for all j, so the second Kuhn–Tucker condition (3) is vacuously true and x^* must be a solution. But the strict convexity of $F(x)$ in the case of $\alpha < 1$ implies a unique solution. Therefore, the sequence has only one subsequential limit point, and so must converge (to the solution, as we have just seen). This is the case considered in [9].

To prove the convergence of $\{x^k\}$ for the case of $\alpha = 1$, Vardi *et al.* [6] use an inequality due essentially to Csiszar and Tusnady [18]; see also Cover [37].

Lemma 4: Let $\alpha = 1$. With x^* as above, $\mathrm{KL}(x^*, x^{k+1}) \leq \mathrm{KL}(x^*, x^k)$ for all $k = 1, 2, \cdots$.

Proof: See [6].

We use Lemma 4 to conclude that $\mathrm{KL}(x^*, x^k)$ goes to zero for the entire sequence, not just for the subsequence. It follows that x^* is unique and that the sequence converges.

For any $x \geq 0$ with $y = Px$, we have that $\mathrm{KL}(x, x^k) \geq \mathrm{KL}(x, x^{k+1})$ since

$$\begin{aligned}\mathrm{KL}(x, x^k) - \mathrm{KL}(x, x^{k+1}) &= \sum x_j \log(x_j^{k+1}/x_j^k)\\ &= -\sum x_j \log\Big(\sum P_{i,j} P x_i^k / y_i\Big)\\ &\geq -\sum x_j \Big(\sum P_{i,j} P x_i^k / y_i - 1\Big)\\ &= \sum P x_i^k - \sum P x_i = 0.\end{aligned}$$

Therefore, $\mathrm{KL}(x, x^\infty) \leq \mathrm{KL}(x, x^k)$ for all k and for all $x \geq 0$ with $y = Px$; it follows that $\mathrm{KL}(x, x^\infty)$ is finite. If $x^\#$ is such that $y = Px^\#$ and $\mathrm{KL}(x, x^\#) \leq \mathrm{KL}(x, x^k)$ for all x with $y = Px$ and all k, then $\mathrm{KL}(x^\infty, x^\#) \leq \mathrm{KL}(x^\infty, x^k)$ for all k, so that $\mathrm{KL}(x^\infty, x^\#) \leq \mathrm{KL}(x^\infty, x^\infty) = 0$; therefore, $x^\infty = x^\#$.

We summarize these results in Theorem 1.

Theorem 1: The sequence $\{x^k\}$ converges to a limit x^∞ for all I and J, for all starting points $x^0 > 0$, for all $0 \leq \alpha \leq 1$, and for all $p > 0$. For $0 \leq \alpha < 1$, x^∞ is the unique minimizer of $F(x)$. For $\alpha = 1$, x^∞ is the unique nonnegative minimizer of $\mathrm{KL}(y, Px)$ if there is no nonnegative solution of $y = Px$. If there are nonnegative solutions, then $y = Px^\infty$ and $x^\# = x^\infty$ is the only solution for which the inequalities $\mathrm{KL}(x, x^\#) \leq \mathrm{KL}(x, x^k)$ hold for all $x \geq 0$ with $y = Px$ and all k; it follows that support(x) is contained in support(x^∞) for all such x.

In the case of $\alpha = 1$, the limit x^∞ of the sequence will be independent of x^0 if there is no nonnegative solution of $y = Px$ (the inconsistent case). If there is a nonnegative solution of $y = Px$, however, x^∞ is also a solution, but may depend on x^0; how it depends on x^0 appears to be an open question. The obvious conjecture, in light of Theorem 2 below, is that x^∞ is the unique solution minimizing $\mathrm{KL}(x^0, x)$; there are simple counterexamples, however.

VI. The Iterative Method for Problem B

We now consider the iterative scheme derived from the two steps above, with $d(r, q) = G(q, r)$. Performing Step 1, we obtain $r^m = r(x^m)$ as before; note that in Problem B, we shall use the index m to denote the iteration number in place of k, simply to help distinguish the two iterative schemes. Performing Step 2, we obtain

$$0 = \alpha\Big[\sum P_{i,j} \log(q_{i,j}^{m+1}/r_{i,j}^{m+1})] + (1-\alpha)[\log(x_j^{m+1}/p_j)\Big] \tag{10}$$

with the sum on i. Solving for x_j^{m+1}, we get

$$x_j^{m+1} = (x_j^m)^\alpha (p_j)^{1-\alpha} \exp\Big(\alpha \sum P_{i,j} \log(y_i/Px_i^m)\Big). \tag{11}$$

For $\alpha = 1$, this is closely related to the multiplicative algebraic reconstruction technique (MART) of Gordon *et al.* [32]; see Section VII below. Because MART varies the factors depending on i, while (11) computes the factors simultaneously, using the current x^m, we shall call (11) the "simultaneous" MART (SMART) algorithm. Lent [35], [46] 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 [33], Censor presents a regularized version of MART, and remarks that the behavior of MART in the inconsistent case is unknown. In [39], Herman notes that in the inconsistent case, MART may not maximize entropy. As we shall show, the SMART iteration scheme in (11) converges to the unique minimizer of $G(x)$ for $\alpha < 1$. For $\alpha = 1$, SMART converges to the unique minimizer of $\mathrm{KL}(Px, y)$ in the inconsistent case; in the consistent case, both SMART and MART converge to the unique nonnegative solution of $y = Px$ minimizing $\mathrm{KL}(x, x^0)$. As with the previous problem, orthogonality conditions of a Pythagorean type will play an important role in the proof of convergence of $\{x^m\}$.

We now present some useful orthogonality conditions relating to the iterative sequence $\{x^m\}$.

Lemma 5: $G(q(x), r(x^m)) - G(q(x^{m+1}), r(x^m)) = \mathrm{KL} \cdot (q(x), q(x^{m+1})) = \mathrm{KL}(x, x^{m+1})$ for all $x \geq 0$.

Proof: This is a consequence of (10).

Lemma 6: $G(q(x^m), r(x)) - G(q(x^m), r(x^m)) = \alpha[\mathrm{KL} \cdot (x^m, x) - \mathrm{KL}(Px^m, Px)] \geq 0$ for all $x \geq 0$.

Proof: This is a simple calculation. Note that $G(q(x), r(x)) = \mathrm{KL}(Px, y)$ for all x.

Lemma 7: $\mathrm{KL}(x^m, x^{m+1})$ goes to zero as m goes to infinity.

Proof: The proof is similar to that of Lemma 3.

We now prove the convergence of the sequence $\{x^m\}$. First we show that the sequence is contained within a bounded set, so that it has subsequential limit points.

Lemma 8: For $\alpha = 1$, $\sum x_j^m \leq \sum y_i$ for each m. For $\alpha < 1$, we have that, for all m, $\sum x_j^m \leq (\sum p_j)^{1-\alpha}(\sum y_i)^\alpha$. In either case, the sequence $\{x^m\}$ is contained within a bounded subset.

Proof: For $\alpha = 1$, $x_j^{m+1} = x_j^m \exp(\sum P_{i,j} \log(y_i/px_i^m))$, and the concavity of the log function gives $x_j^{m+1} \leq x_j^m \exp(\log(\sum P_{i,j} y_i/Px_i^m))$, from which the first result follows by summing on j. To prove the second assertion, we write $x_j^{m+1} \leq (x_j^m)^\alpha (p_j)^{1-\alpha}[\sum P_{i,j}(y_i/Px_i^m)]^\alpha$ and use Holder's inequality.

Now let x^* be the limit of the subsequence $\{x^{mn}\}$. It follows from Lemma 7, as in the case of Problem A, that x^* is a fixed point of the iterative procedure in (11); therefore, it satisfies the first of the two Kuhn–Tucker conditions (2) for $f(x) = G(x)$.

The Case of $\alpha < 1$*:* In the case of $\alpha < 1$, we show that $x_j^* > 0$ for every j. From this, it follows that the second Kuhn–Tucker condition (3) is vacuously true and that x^* is a solution. We need two lemmas.

Lemma 9: The sequence $\{\sum P_{i,j} \log(y_i/Px_i^m)\}$, with the sum on i, is bounded below for each fixed j.

Proof: For each fixed m, we have $Px_i^m \leq \sum x_j^m \leq \sum y_i = y_+$ from Lemma 8 and our assumption that $\sum y_i = \sum p_j$. Therefore, we have $y_i/Px_i^m \geq y_i/y_+$ and $\log(y_i/Px_i^m) \geq \log(y_i/y_+)$ for each i. Therefore, $\sum P_{i,j} \log(y_i/Px_i^m) \geq \sum P_{i,j} \log(y_i/y_+) = C_j$ for all m and for all j.

Lemma 10: $x_j^* > 0$ for all j.

Proof: From $x_j^{m+1} = (x_j^m)^\alpha (p_j)^{1-\alpha} \exp(\alpha \sum P_{i,j} \log \cdot (y_i/Px_i^m))$, it follows that $\log x_j^{m+1} = \alpha \log(x_j^m) + (1-\alpha)\log(p_j) + (\alpha \sum P_{i,j} \log(y_i/Px_i^m))$. We then obtain $\log(x^{m_n+1_j}) \geq (1-\alpha^{m_n+1}) \log(p_j) + \alpha^{m_n+1} \log(x_j^0) + \alpha(1-\alpha^{m_n+1}/(1-\alpha)C_j$. As m_n goes to infinity, the right-hand side converges to a finite quantity, so the left side cannot converge to $-\infty$.

It follows that x^* is the unique solution and the limit of $\{x^m\}$. This completes the analysis for the case of $\alpha < 1$.

The Case of $\alpha = 1$*:* For the case of $\alpha = 1$, how we proceed depends on whether or not the system of equations $y = Px$ has nonnegative solutions.

a) The Consistent Case: Let x^* be the limit of the subsequence $\{x^{m_n}\}$, and let x be such that $y = Px$. From Lemma 5, we have, for all m,

$$\begin{aligned} \mathrm{KL}(x, x^m) &= \mathrm{KL}(y, Px^m) + \mathrm{KL}(x, x^{m+1}) \\ &\quad + \mathrm{KL}(q^{m+1}, r^{m+1}) \geq \mathrm{KL}(x, x^{m+1}). \end{aligned} \tag{12}$$

Therefore, $\{\mathrm{KL}(x, x^m) - \mathrm{KL}(x, x^{m+1})\}$ converges to zero. But from (12), we have $\mathrm{KL}(x, x^m) - \mathrm{KL}(x, x^{m+1}) \geq \mathrm{KL}(y, Px^m)$. It follows that $\mathrm{KL}(y, Px^*) = 0$ or that $y = Px^*$. Since $\mathrm{KL}(x, x^m) - \mathrm{KL}(x, x^{m+1})$ is independent of x, if $y = Px$, x^* is the unique solution of $y = Px$ minimizing $\mathrm{KL}(x, x^0)$. But since this is true of each x^*, x^* must be unique, and the sequence $\{x^m\}$ must converge to x^*.

b) The Inconsistent Case: Again, let x^* be the limit of the subsequence $\{x^{m_n}\}$. We know that x^* is a fixed point of the iteration. Therefore, we have from Lemma 5 that, for any $x \geq 0$,

$$\begin{aligned} \mathrm{KL}(q(x), r(x^*)) &= \mathrm{KL}(q(x^*), r(x^*)) + \mathrm{KL}(x, x^*) \\ &= \mathrm{KL}(Px^*, y) + \mathrm{KL}(x, x^*), \end{aligned} \tag{13}$$

and from Lemma 6,

$$\begin{aligned} \mathrm{KL}(q(x), r(x^*)) &= \mathrm{KL}(q(x), r(x)) + \mathrm{KL}(x, x^*) \\ &\quad - \mathrm{KL}(Px, Px^*) \\ &= \mathrm{KL}(Px, y) + \mathrm{KL}(x, x^*) - \mathrm{KL}(Px, Px^*). \end{aligned} \tag{14}$$

Therefore, we have $\mathrm{KL}(Px^*, y) \leq \mathrm{KL}(Px, y)$; it follows that x^* is a global minimizer of $\mathrm{KL}(Px, y)$. But by Proposition 2 below, the minimizer is unique in the inconsistent case. Therefore, the x^* is unique, and the sequence $\{x^m\}$ must converge to x^*.

Proposition 2: Let $J \geq I$, and let $M = \{$all $x \geq 0$ for which $\mathrm{KL}(Px, y)$ attains its minimum value$\}$. Then either

1) the minimum of $\mathrm{KL}(Px, y)$ is zero and $y = Px$ has nonnegative solutions, or

2) there is a subset S of $\{1, 2, \cdots, J\}$ having I-1 or fewer members such that if x is in M, then $x_j > 0$ only if j is in S.

In the former case, there may be multiple solutions, so M may contain more than one vector. In the latter case, M contains only one vector, the unique minimizer of $\mathrm{KL}(Px, y)$.

Proof: The first Kuhn–Tucker condition for a minimizer of $\mathrm{KL}(Px, y)$ is now $0 = \sum P_{i,j} \log(y_i/Px_i)$ for each j such that $x_j > 0$. Define Q as in the proof of Proposition 1 and the vector $v = (v_i) = (\log(y_i/Px_i))$. Then we have $Q^T v = 0$. If $K \geq I$, then $v = 0$, so that $y = Px$. The assertion follows as in the earlier proposition.

We summarize these results in Theorem 2.

Theorem 2: The sequence $\{x^m\}$ converges to a limit x^∞ for all I and J, for all starting points $x^0 > 0$, for all $p > 0$, and for all $0 \leq \alpha \leq 1$. For $0 \leq \alpha < 1$, x^∞ is the unique minimizer of $G(x)$. For $\alpha = 1$, x^∞ is the unique nonnegative minimizer of $\mathrm{KL}(Px, y)$ if there is no nonnegative solution of $y = Px$. If there are nonnegative solutions of $y = Px$, then the limit may depend on the starting value; we have $y = Px^\infty$, x^∞ is the unique solution minimizing $\mathrm{KL}(x, x^0)$, and support(x) is contained within support(x^∞) for all $x \geq 0$ with $y = Px$.

VII. Comparing the SMART and MART Algorithms

In this section, we give the MART algorithm of Gordon *et al.* [32] and compare its convergence with that of SMART; details concerning the convergence of MART will be presented separately [41].

The MART iterative procedure is the following: starting with $z^0 = x^0 > 0$, let

$$z_j^{k+1} = z_j^k (y_i/Pz_i^k)^{P^{ij}} \tag{15}$$

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

$$x_j^{n+1} = x_j^n \prod (y_i/Pz_i^{nI+i-1})^{P^{i,j}} \tag{16}$$

with the product over i. Compare MART with SMART iteration (for $\alpha = 1$) defined by

$$x_j^{m+1} = x_j^m \prod (y_i/Px_i^m)^{P^{i,j}}; \tag{17}$$

the difference between x^n and x^m is that, in the computation of the latter, the factors in the product are not updated one at a time as i changes, but simultaneously at the end of each cycle, hence the name "simultaneous MART" (SMART). This makes SMART more "parallelizable."

To place MART within a framework of alternating projections, we define, for each i and for each x and z, the distance $G_i(q(x), r(z))$:

$$\begin{aligned} G_i(q(x), r(z)) = \sum [&q(x)_{i,j} \log(q(x)_{i,j}/r(z)_{i,j}) \\ &+ r(z)_{i,j} - q(x)_{i,j}] \end{aligned}$$

$$+\sum (1-P_{i,j})[x_j \log(x_j/z_j) + z_j - x_j] \tag{18}$$

with both sums over j only. The MART iterative procedure can then be obtained as follows: beginning with $z^0 = x^0 > 0$, for each $k = 1, 2, \cdots$ and with $i = i(k) = k \bmod I + 1$, perform the following alternating minimizations.

Step 1(i): Minimize $\mathrm{KL}(q(z^k), r(z))$ to get $r^{k+1} = r(z^k)$.

Step 2(i): Minimize $G_i(q(x), r(z^k)$ to get $x = z^{k+1}$.

In the consistent case, we have a short proof [41] that the MART sequence $\{z^k\}$ converges to the same limit as SMART, namely, the unique nonnegative solution of $y = Px$ for which $\mathrm{KL}(x, x^0)$ is minimized. For the case of a uniform starting vector x^0, the limit will then be the maximum Shannon entropy solution, as Lent has shown [35], [46].

In [33], Censor remarks that the behavior of MART in the inconsistent case is not known. In small-dimensional simulation studies of MART in the inconsistent case, we have observed some behavior patterns which we now conjecture hold more generally. It is easy to show that the full MART sequence $\{z^k\}$ cannot converge in the inconsistent case; however, in our simulations, the subsequences $\{z^{nI+i}, i \text{ fixed}\}$ associated with completed cycles converge to distinct limit vectors $z^{\infty,i}$. The support of each $z^{\infty,i}$ is that of the SMART limit x^∞, and the final projection data $\{Pz_i^{\infty,i}\}$ is quite close to the initial projection data $\{y_i\}$ in the sense that $\mathrm{KL}(\{Pz_i^{\infty,i}\}, \{y_i\})$ is small. If we use this final projection data $\{Pz_i^{\infty,i}\}$ in place of $\{y_i\}$ as initial projection data and repeat the MART iteration process, we find that the limits $\{z^{\infty,i}\}$, although still distinct, are now closer to one another than previously, and the new final and initial projection data are closer than before. Repeating this feedback into the MART algorithm several times, we see the distinct limits $\{z^{\infty,i}\}$ growing closer to each other and converging to the SMART limit x^∞. We hope to present proof shortly that these behavior patterns are generally followed [41].

Censor and his colleagues have recently introduced block-iterative methods and have studied their implementation [43]–[46]. The MART algorithm is one special case, in which each equation has its own block; at the other extreme is the algorithm in which all equations are dealt with in a single block. The latter algorithm is nearly the unrelaxed SMART, differing from it only by the presence in their algorithm of a weighting; the weighting appears unnecessary and probably slows convergence. They do not give convergence proofs for the inconsistent case for these more general algorithms, nor do they place them within a framework of alternating optimization.

VIII. Summary and Conclusions

We consider here the related Problems A and B of minimizing the functionals $F(x) = \alpha\mathrm{KL}(y, Px) + (1-\alpha)\mathrm{KL}(p, x)$ and $G(x) = \alpha\mathrm{KL}(Px, y) + (1-\alpha)\mathrm{KL}(x, p)$, respectively, over the set of vectors $x \geq 0$. We have derived iterative algorithms for minimizing both functionals using the method of alternating projections, extending the approach used by Csiszar and Tusnady [18] and Vardi *et al.* [6] to demonstrate the convergence of the EM/ML algorithm (equivalently, Problem A with $\alpha = 1$). For Problem A and $\alpha < 1$, we obtain the iterative MAP method of Lange *et al.* [9].

For Problem B, we present a "simultaneous" version of the MART algorithm of Gordon *et al.* [32], called SMART, and prove the convergence of SMART. In the consistent case, both MART and SMART converge to the unique solution minimizing $\mathrm{KL}(x, x^0)$; in the inconsistent case, SMART converges to the unique nonnegative minimizer of $\mathrm{KL}(Px, y)$. We also consider a regularized SMART algorithm and prove convergence to the unique minimizer of $G(x)$.

It is claimed in [6] that whenever the number of unknowns is greater than the number of data values, there will be nonunique minima of $\mathrm{KL}(y, Px)$; on the contrary, there are unique nonnegative minimizers of $\mathrm{KL}(y, Px)$ and $\mathrm{KL}(Px, y)$ unless the system of equations $y = Px$ has a nonnegative solution. When $y = Px$ has nonnegative solutions and $\alpha = 1$, the limit of the iterations x^∞ may depend on the starting value x^0. For Problem B, we have that x^∞ is the unique solution minimizing $\mathrm{KL}(x, x^0)$; the corresponding question for Problem A appears to be open.

Acknowledgment

The author would like to thank Dr. J. Graham-Eagle of UML for his many contributions to the present paper, as well as colleagues at the Department of Nuclear Medicine, University of Massachusetts Medical Center, for their continued encouragement and assistance; particular thanks go to Dr. M. King, Dr. B. Penney, Dr. S. Glick, and Dr. N. Rajeevan, as well as to Dr. M. Gennert of Worcester Polytechnic Institute and P. Sutter of ETH, Zurich. The author also thanks Prof. Y. Censor of the University of Haifa for bringing to his attention [43]–[46], and the anonymous reviewer for a copy of [42]. Equation (11) with α=1 is also the iterative scaling algorithm of Darroch and Ratcliff, *Annals of Stat.*, vol 17, no. 3, pp. 1409–1413, 1989.

References

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

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

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

[4] Z. Liang and H. Hart, "Bayesian image processing of data from constrained source distributions—I. Non-valued, correlated and uncorrelated constraints," *Bull. Math. Biol.,* vol. 49, no. 1, pp. 51–74, 1987.

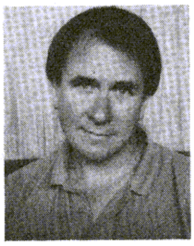
[5] D. Snyder and M. Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," *IEEE Trans. Nucl. Sci.,* vol. NS-32, pp. 3864–3872, Oct. 1985.

[6] Y. Vardi, L. Shepp, and L. Kaufman, "A statistical model for positron emission tomography, *J. Amer. Statist. Ass.,* vol. 80, pp. 8–37, 1985.

[7] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imaging,* vol. 8, pp. 194–202, June 1989.

[8] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-6, pp. 721–741, Nov. 1984.

[9] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Trans. Med. Imaging,* vol. MI-6, pp. 106–114, June 1987.
[10] E. Levitan and G. Herman, "A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. Med. Imaging,* vol. MI-6, pp. 185–192, Sept. 1987.
[11] Z. Liang, R. Jaszczak, C. Floyd, and K. Greer, "A spatial interaction model for statistical image processing," presented at the 11th Int. Conf. Inform. Processing in Med. Imaging, Berkeley, CA, 1989.
[12] P. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imaging,* vol. 9, pp. 84–93, Mar. 1990.
[13] S. Geman and D. McClure, "Bayesian image analysis: An application to single photon emission tomography," in *Proc. Statist. Comput. Sect.,* Amer. Statist. Assoc., Washington, DC, 1985, pp. 12–18.
[14] R. Leahy, T. Hebert, and R. Lee, "Applications of Markov random field models in medical imaging," in *Proc. Conf. Inform. Processing in Med. Imaging, IPMI-89,* Lawrence–Berkeley Lab., June 1989.
[15] B.W. Silverman, M.C. Jones, J.D. Wilson, and D.W. Nychka, "A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography," *J. Roy Statist. Soc. B,* vol. 52, no. 2, pp. 271–324, 1990.
[16] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Statist.,* vol. 22, no. 1, pp. 79–86, 1951.
[17] D. Titterington, "On the iterative image space reconstruction algorithm for ECT," *IEEE Trans. Med. Imaging,* vol. MI-6, pp. 52–56, Mar. 1987.
[18] I. Csiszar and G. Tusnady, "Information geometry and alternating minimization procedures," *Statist. Decisions,* suppl. issue no. 1, pp. 205–237, 1984.
[19] C.R. Smith and G.J. Erickson, Ed., *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems.* Reidel, 1987.
[20] R. MacKinnon, "Minimum cross-entropy noise reduction in images," *J. Opt. Soc. Amer A.,* vol. 6, pp. 739–747, June 1989.
[21] E. Jaynes, "On the rationale of maximum-entropy methods," *Proc. IEEE,* vol. 70, pp. 939–952, Sept. 1982.
[22] I. Csiszar, "Why least squares and maximum entropy? An axiomatic approach to inverse problems," preprint 19/1989, Math. Inst. of the Hungarian Acad. of Sci., Budapest.
[23] J. Shore and R. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory,* vol. IT-26, pp. 26–37, Jan. 1980.
[24] J. Shore, "Minimum cross-entropy spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-29, pp. 230–237, Apr. 1981.
[25] L. Jones, "Approximation theoretic derivation of logarithmic entropy principles for inverse problems and unique extension of the maximum entropy method to incorporate prior knowledge," *SIAM J. Appl. Math.,* vol. 49, pp. 650–661, Apr. 1989.
[26] L. Jones and C. Byrne, "General entropy criteria for inverse problems, with applications to data compression, pattern classification and cluster analysis," *IEEE Trans. Inform. Theory,* vol. 36, pp. 23–30, Jan. 1990.
[27] C. Byrne and L. Jones, "An axiomatic approach to certain inverse problems," in *SPIE Conf. Proc.,* vol. 1351, San Diego, CA, July 1990.
[28] J. Burg, "The relationship between maximum entropy spectra and maximum likelihood spectra," *Geophys.,* vol. 37, pp. 375–376, Apr. 1975.
[29] R. Lacoss, "Data adaptive spectral analysis methods," *Geophys.,* vol. 36, pp. 671–675, Aug. 1971.
[30] B. Frieden, "Restoring with maximum likelihood and maximum entropy," *J. Opt. Soc. Amer.,* vol. 62, pp. 511–518, Apr. 1972.
[31] S. Gull and G. Daniell, "Image reconstruction from incomplete and noisy data," *Nature,* vol. 272, pp. 666–670, Apr. 1978.
[32] 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.
[33] Y. Censor, "Finite series-expansion reconstruction methods," *Proc. IEEE,* vol. 71, pp. 409–419, Mar. 1983.
[34] Y. Censor, T. Elfving, and G. Herman, "Special-purpose algorithms for linearly constrained entropy maximization," in *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems,* C.R. Smith and G.J. Erickson, Ed. Reidel, 1987, pp. 241–254.
[35] 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, Washington, DC, 1977, pp. 238–243 (also [46]).
[36] A. Peressini, F. Sullivan, and J. Uhl, *The Mathematics of Nonlinear Programming.* New York: Springer-Verlag, 1988.
[37] T. Cover, "An algorithm for maximizing log investment return," *IEEE Trans. Inform. Theory,* vol. IT-30, pp. 369–373, Mar. 1984.
[38] C.P. Smith and W.T. Grandy, Jr., Ed. *Maximum Entropy and Bayesian Methods in Inverse Problems.* Reidel, 1985.
[39] G. Herman, "Applications of maximum entropy and Bayesian optimization methods to image reconstruction from projections," in *Maximum Entropy and Bayesian Methods in Inverse Problems,* C.P. Smith and W.T. Grandy, Jr., Ed. Reidel, 1985, pp. 319–338.
[40] W.H. Fleming, *Functions of Several Variables.* Reading, MA: Addison-Wesley, 1965.
[41] C. Byrne and J. Graham-Eagle, "Convergence properties of the multiplicative algebraic reconstruction technique (MART)," in progress.
[42] L. Shepp and R. Vanderbei, "New insights into emission tomography via linear programming," AT&T Tech. Memo., Nov. 1988.
[43] Y. Censor and J. Segman, "On block-iterative maximization," *J. Inform. Optimiz. Sci.,* vol. 8, no. 3, pp. 275–291, 1987.
[44] S. Zenios and Y. Censor, "Parallel computing the block-iterative image reconstruction algorithms," *Appl. Numer. Math.,* vol. 7, pp. 399–415, 1991.
[45] Y. Censor, "Parallel application of block-iterative methods in medical imaging and radiation therapy," *Math. Programming,* vol. 42, pp. 307–325, 1988.
[46] A. Lent and Y. Censor, "The primal-dual algorithm as a constraint-set-manipulation device," *Math. Programming,* vol. 50, pp. 343–357, 1991.



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 Professor in the Department of Mathematics, The Catholic University of America, Washington, DC, serving as Chairman from 1983 to 1986. Since 1986 he has been in the Department of Mathematics, University of Massachusetts at Lowell, serving as Chairman from 1987 to 1990. He has been a consultant in acoustic signal processing to the Naval Research Laboratory, the Office of Naval Research, and the Australian Department of Defence. Since 1990 he has been a consultant in tomographic image processing to the Department of Nuclear Medicine, University of Massachusetts Medical Center, Worcester, MA.

SEE BYRNE, TRANS IP, VOL 4, NO. 2, FEB. 1995