Likelihood Maximization for List-Mode Emission Tomographic Image Reconstruction

Charles Byrne, *Member, IEEE*

Abstract—**The maximum *a posteriori* (MAP) Bayesian iterative algorithm using priors that are gamma distributed, due to Lange, Bahn and Little, is extended to include parameter choices that fall outside the gamma distribution model. Special cases of the resulting iterative method include the *expectation maximization maximum likelihood (EMML)* method based on the Poisson model in emission tomography, as well as algorithms obtained by Parra and Barrett and by Huesman *et al.* that converge to maximum likelihood and maximum conditional likelihood estimates of radionuclide intensities for list-mode emission tomography. The approach taken here is optimization-theoretic and does not rely on the usual *expectation maximization* (EM) formalism. Block-iterative variants of the algorithms are presented. A self-contained, elementary proof of convergence of the algorithm is included.**

Index Terms—**Emission tomography, image reconstruction, iterative algorithm, probabilistic mixture.**

I. Introduction

In *bin-mode* single-photon emission computed tomography (SPECT) photons are detected at one of I predetermined detector locations. During the course of the scanning, most detector locations count multiple arrivals, the counts at each detector are modeled as independent Poisson random variables, and likelihood maximization is used to estimate the radionuclide intensity levels at each voxel [29], [19].

For positron emission tomography (PET) coincidence imaging, a detection corresponds to the (nearly) simultaneous arrival of two photons at opposite ends of a line segment, or line of response, joining a pair of detector locations. The set of all such pairs is quite large, and in the course of the scanning, most of these pairs are associated with no detected events, some with one and typically none with more than one. The bin-mode approach can introduce distortions in the imaging by lumping together detected events that differ significantly, thereby ignoring potentially useful information. Because it is highly inefficient to store a large number of zero values and a smaller number of values that are one, it is common to use *list-mode* processing, whereby one maintains a list of those pairs associated with a detected event, possibly including additional information, such as the energies of the detected pairs of photons. Once again, likelihood maximization is used to estimate the radionuclide intensities at the voxels. Now, however, the Poisson model is inappropriate to describe the list and the likelihood function is that of a probabilistic mixture [27].

In both of these situations, the *expectation maximization* (EM) method [13] can be used to derive an iterative procedure to maximize the likelihood, as was done for bin-mode by Shepp and Vardi [29] and Lange and Carson [19], and for list-mode by Parra and Barrett and by Huesman *et al.* [16]. In these papers, the EM formalism was employed to obtain an iterative method for likelihood maximization.

In this paper, we avoid the EM approach by rewriting the likelihood function as a penalized cross-entropy and applying our algorithm for minimizing such functions. Our algorithm includes, as particular cases, the EM algorithm for SPECT of Shepp and Vardi [29] and Lange and Carson [19], called here the EMML algorithm, the Bayesian maximum *a posteriori* (MAP) method of Lange *et al.* [20], and the likelihood maximization method of Parra and Barrett [26] and of Huesman *et al.* [16]. In fact, our iterative algorithm is an extension of the MAP method for regularized likelihood maximization given in [20], involving gamma-distributed prior distributions on the voxel intensities. While formally identical to their MAP method, ours allows the parameters to take on nonpositive values, which are outside the gamma distribution model. With this added flexibility, we obtain the Parra and Barrett algorithm as a particular case. Negative parameter values can arise when we use probability density functions (pdf's) to describe the distribution of possible events, rather than probability functions (pf's).

The Poisson model is used only to describe the total counts, as in [26]. The bin-mode EM algorithm for SPECT can be rederived in this way, without the additional assumptions that the individual counts at each detector are Poisson distributed. Our approach is entirely optimization-theoretic and does not involve the EM formalism.

In Section II, we give a short history of recent efforts to prove convergence of the EM algorithm and its various relatives, both to motivate the inclusion of yet another proof here, and to put our proof in some context. List-mode emission tomography and likelihood maximization are treated in Section III. In Section IV, we present a brief overview of the iterative algorithms we shall be discussing. By viewing them together we can better see the similarities and subtle differences that will be important later in the discussion. Section V concerns our new iterative algorithm, with particular cases of the algorithm discussed in Section VI. In Section VII, we look at the relationship between the iterative EM algorithm for list-mode processing and the EMML

Manuscript received April 30, 2001; revised August 6, 2001. This work was supported in part by the National Institutes of Health (NIH) under Grant CA23452 to the University of Massachusetts Medical School. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was J. Liang.

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

Publisher Item Identifier S 0278-0062(01)09310-7.

bin-mode algorithm. A block-iterative version of our algorithm is presented in Section VIII. Underlying the formulation of the likelihood function for the list-mode case is the notion of a probabilistic mixture; in Section IX we give a simple example of such a mixture and show how our algorithm can be applied to estimate the mixing coefficients. A self-contained and elementary proof of convergence for our algorithm is given in the Appendix.

II. A Brief History of Convergence Proofs for the EM Algorithm

It seems customary to credit the EM algorithm to Dempster, Laird and Rubin [13], although it is made clear in that paper that the EM algorithm did not originate with them. For detailed histories of the EM algorithm, see the paper by Meng and van Dyk [24] and the recent text by McLachlan and Krishnan [23].

The EM algorithm concerns a parameterized family of pdf's or pf's $\{f(\cdot|x)\}$, where x is a parameter vector. The algorithm generates a sequence $\{x^k\}$ that we hope will converge to a parameter vector maximizing the likelihood. The tricky point is to relate convergence of the sequence of functions $\{f(\cdot|x^k)\}$ to convergence of the sequence $\{x^k\}$. The convergence proof in [13] contained an error involving just this point. Wu [31] pointed out this error and provided sufficient conditions for convergence, in terms of a forcing function relating the behavior of the $f(\cdot|x)$ to that of the x (see also Redner and Walker [27]).

In [29], Shepp and Vardi consider the special case of the EM applied to likelihood maximization in emission tomography; we call this special case the EMML algorithm in this paper. As Lange and Carson note in [19], the convergence proof in [29] has a gap, in that it ignores the possibility that the limit vector of the iterative sequence may have zero entries. The proof given by Lange and Carson is valid, but requires the additional assumption that the likelihood maximizer be unique. In [30], the authors acknowledge the point made by Lange and Carson and correct the earlier proof using results of Csiszár and Tusnády [12] and Cover [11].

There is a misstatement in the theorem in [30], however. They assert (incorrectly) that the likelihood has a unique maximizer if and only if the rows of the matrix of probabilities involved span R^J, where J is the number of voxels; for the bin-mode case this implies that there are at least as many bins as there are voxels or pixels. If they were correct, the typically noisy maximum likelihood images could be smoothed by using more voxels than bins, to guarantee nonunique likelihood maximizers, generating a number of distinct noisy maximum likelihood images from different starting points, and then averaging them, to obtain a smoother maximum likelihood image. However, as shown in [4], smooth maximum likelihood images cannot be obtained in this way, since the maximum likelihood estimate is almost always unique for noisy data, hence independent of starting point, even when there are more voxels than bins (see also [5]).

To obtain smoother images, regularization is often used. An example is the maximum *a posteriori* Bayesian approach of Lange *et al.* [20], involving prior gamma distributions for the unknowns. Since the mean and variance of the gamma distribution are positive, they do not have any problem with limit vectors having any zero entries.

As noted in [30], the article [12] of Csiszár and Tusnády can be difficult to follow. A short proof of convergence for the EMML algorithm was given by Iusem in [17], using the joint convexity of the KL distance. A more general result along the same lines is given in [9]. A shorter and more elementary proof of convergence of the EMML was given in [6]. The convergence proof given in this paper is an extension of the proof for EMML in [6].

III. Overview of the Relevant Algorithms

Central to our discussion is the notion of cross-entropy or Kullback–Leibler (KL) distance [18] between vectors with nonnegative entries. For $a > 0$ and $b > 0$, let $\mathrm{KL}(a, b) = a\log(a/b) + b - a$, $\mathrm{KL}(a, 0) = +\infty$ and $\mathrm{KL}(0, b) = b$. For nonnegative M-dimensional vectors w and v, let $\mathrm{KL}(w, v)$ be defined component-wise as $\mathrm{KL}(w, v) = \sum_{m=1}^{M} \mathrm{KL}(w_m, v_m)$.

The EMML algorithm for SPECT [29], [19] can be viewed as an iterative procedure for finding a nonnegative (approximate) solution x for a system of linear equations of the form $y = Px$, where $y = (y_1, \ldots, y_N)^T$ is a vector with positive entries and P is a matrix with entries $P_{nj} \geq 0$, for $n = 1, \ldots, N$ and $j = 1, \ldots, J$, such that $s_j = \sum_{n=1}^{N} P_{nj} > 0$ for each j. The EMML algorithm converges to a minimizer of the function $\mathrm{KL}(y, Px)$ over all nonnegative vectors. With $(Px)_n = \sum_{j=1}^{J} P_{nj}x_j$, the iterative step of the EMML is

The EMML:

$$x_j^{k+1} = s_j^{-1}x_j^k \sum_{n=1}^{N} P_{nj}\frac{y_n}{(Px^k)_n}. \tag{3.1}$$

When the EMML algorithm is used to reconstruct images from noisy data, the result is often too rough to be useful. Variations of the EMML involving regularization are then employed to obtain smoother images.

One such regularization method is the maximum *a posteriori* (MAP) Bayesian method of Lange *et al.* [20], in which each unknown entry x_j is assumed to have a prior gamma distribution, with mean $\mu_j > 0$ and variance σ_j^2 assumed to lie within the interval $(0, \mu_j)$. The function to be minimized can be written as

$$G(x) = \mathrm{KL}(y, Px) + \sum_{j=1}^{J} \beta_j \mathrm{KL}(\gamma_j, x_j) \tag{3.2}$$

where $\beta_j = \mu_j/\sigma_j^2 > 0$ and $\gamma_j = (\mu_j - \sigma_j^2)/\mu_j > 0$. The iterative step of their MAP method is

The MAP Method With Gamma Priors:

$$x_j^{k+1} = \alpha_j s_j^{-1}x_j^k \sum_{n=1}^{N} P_{nj}\frac{y_n}{(Px^k)_n} + (1 - \alpha_j)\gamma_j \tag{3.3}$$

with $\alpha_j = s_j/(s_j + \beta_j)$. Note that if we violate the restrictions on the parameters γ_j and take $\gamma_j = 0$ for each j, then the function to be minimized becomes

$$G(x) = \mathrm{KL}(y, Px) + \sum_{j=1}^{J} \beta_j x_j. \tag{3.4}$$

As we shall see, maximizing the likelihood in list-mode image processing in SPECT and PET is equivalent to minimizing a function of the form in (3.4).

In the list-mode case, the vector y is the vector u whose entries are all one and the β_j may sometimes be negative, which is not permitted in the gamma distribution. The iterative algorithm for this case has the following update equation

The List-Mode EM Algorithm:

$$x_j^{k+1} = d_j^{-1} x_j^k \sum_{n=1}^{N} P_{nj} \frac{1}{(Px^k)_n} \tag{3.5}$$

with $d_j \in (0, 1]$ denoting the probability that an emission at the jth voxel will be detected, and $\beta_j = d_j - s_j$. This is the algorithm for list-mode processing given by Parra and Barrett [26] and by Huesman *et al.* [16]. When pdf's are used to describe the probability distribution of events, the entries of the matrix P are values of the pdf, so the s_j can be any positive numbers, while the d_j are restricted to $(0, 1]$.

The main goal in this paper is to derive and prove convergence of an extension of the MAP method in (3.3) that is sufficiently general to include the list-mode EM algorithm. Specifically, we shall prove that the iterative algorithm with the update equation.

Our Algorithm:

$$x_j^{k+1} = \alpha_j s_j^{-1} x_j^k \sum_{n=1}^{N} P_{nj} \frac{y_n}{(Px^k)_n} + (1 - \alpha_j)\gamma_j \tag{3.6}$$

with $\alpha_j = s_j/(s_j + \beta_j)$, converges to a minimizer of the function

$$G(x) = \mathrm{KL}(y, Px) + \sum_{j=1}^{J} \beta_j \mathrm{KL}(\gamma_j, x_j), \tag{3.7}$$

provided $s_j + \beta_j > 0$ and $\beta_j \gamma_j \geq 0$ for all j.

IV. List-Mode Processing

In the list-mode approach, the set of possible outcomes is usually either infinite or so large that typically no single outcome occurs more than once and most never occur at all. Of course, we can, if we wish, consider bin-mode data in list-mode format; in that case, the set of possible outcomes is much smaller and most of them occur many times.

For $j = 1, 2, \ldots, J$ let $x_j \geq 0$, be the expected number of events originating within the jth voxel during the scanning time; we assume that x_j is proportional to the intensity of the radionuclide within the jth voxel. Let $f_j(\cdot)$ the pdf or pf describing the distribution of events originating in the jth voxel. With $x_+ = \sum_{j=1}^{J} x_j$, the probability that an event originated within the jth voxel, given that it originated within one of the J voxels, is x_j/x_+. Therefore, by the familiar rule of total probability, the distribution of random events is described by

$$f(\cdot) = \frac{1}{x_+} \sum_{j=1}^{J} x_j f_j(\cdot). \tag{4.1}$$

This indicates that the distribution of random events can be considered as a probabilistic mixture [27], with the mixing coefficients x_j/x_+ determined by the relative intensities of the radionuclides at each voxel. In a later section, we consider such mixtures in more detail.

We denote by $\{z_1, \ldots, z_N\}$ the list of detected events. We assume that the items on the list were independently generated. The number N of events in the list is also a random quantity; to determine its distribution we need to discuss detectability, that is, the probability that an event originating within one of the voxels will be detected.

Let the set of possible events be denoted by Z. If Z is an infinite set or a finite set so large that typically no single outcome occurs more than once and most never occur at all, the functions $f_j(\cdot)$ are pdf or pf defined over Z. We then let $g(z) \in [0, 1]$ be the probability that $z \in Z$ will be detected if it occurs. The probability of detecting an event originating within the jth voxel is assumed to be positive and given by $d_j = \int f_j(z)g(z)\,dz$ or $d_j = \sum f_j(z)g(z)$, depending on whether $f_j(\cdot)$ is a pdf or a pf. Note that d_j does not depend on which events happen to be on the list.

For each j, let d_j in $(0, 1]$ be the probability of detecting an event originating within the jth voxel. Then d, the probability of detecting an event, is given by

$$d = \frac{1}{x_+} \sum_{j=1}^{J} x_j\, d_j. \tag{4.2}$$

We model the random number of items in the list as a Poisson random variable with mean $x_+ d$. Therefore, the probability of having N items in the list is

$$p(N) = \exp(-x_+ d)\,(x_+ d)^N / N!. \tag{4.3}$$

With $x = (x_1, x_2, \ldots, x_J)^T$, the likelihood function $L(x)$ is

$$L(x) = p(N) \prod_{n=1}^{N} f(z_n). \tag{4.4}$$

Taking logs of both sides and using (4.1), we get the log likelihood function, $LL(x)$, given, except for constants, by

$$LL(x) = -x_+ d + N \log(x_+ d) + \sum_{n=1}^{N} \log(Px)_n \tag{4.5}$$

where P is the matrix with entries $P_{nj} = f_j(z_n)$ and $(Px)_n = \sum_{j=1}^{J} P_{nj} x_j$. Note that $\sum_{n=1}^{N} (Px)_n = \sum_{n=1}^{N} \sum_{j=1}^{J} P_{nj} x_j = \sum_{j=1}^{J} (\sum_{n=1}^{N} P_{nj}) x_j = \sum_{j=1}^{J} s_j x_j$.

To estimate the intensity vector x we maximize the function $LL(x)$, subject to $x_j \geq 0$ for each j. To obtain an iterative method for doing this, we rewrite the problem using the KL distance between nonnegative vectors.

With u the vector whose entries are all one, we have

$$-LL(x) = \mathrm{KL}(u, Px) - \sum_{n=1}^{N} (Px)_n + x_+ d + \text{constants}. \tag{4.6}$$

With $s_j = \sum_{n=1}^{N} P_{nj} = \sum_{n=1}^{N} f_j(z_n)$, our problem is to minimize the function

$$F(x) = \mathrm{KL}(u, Px) + \sum_{j=1}^{J} (d_j - s_j) x_j. \tag{4.7}$$

Note that since the functions $f_j(\cdot)$ may be pdf, s_j can be any positive number, so $d_j - s_j$ need not be positive.

The function $F(x)$ that we wish to minimize, over all x with $x_j \geq 0$, is a particular case of the function $G(x)$ given by

$$G(x) = \mathrm{KL}(y, Px) + \sum_{j=1}^{J} \beta_j \mathrm{KL}(\gamma_j, x_j) \tag{4.8}$$

where $y = (y_1, \ldots, y_N)^T$ is a vector with positive entries and $\gamma_j \geq 0$ for each j. To obtain $F(x)$ from $G(x)$ we set $\beta_j = d_j - s_j$ and $\gamma_j = 0$ for each j and $y_n = u_n = 1$ for each n. In the Section Vwe derive an iterative algorithm to minimize $G(x)$ over all x with nonnegative entries. The maximum likelihood algorithm for list-mode processing will then emerge as a particular case.

V. The Iterative Algorithm

Throughout this section we shall assume that y is an N-dimensional vector with positive entries y_n, $n = 1, \ldots, N$ and let $y_+ = \sum_{n=1}^{N} y_n$. We also assume that P is an N by J matrix with nonnegative entries, such that each row and column of P has at least one positive entry. For $j = 1, 2, \ldots, J$, let $s_j = \sum_{n=1}^{N} P_{nj} > 0$ and choose $\gamma_j \geq 0$. For nonnegative vectors x in the set $X = \{x | (Px)_n > 0, n = 1, \ldots, N\}$, let $G(x)$ be the function defined by (4.8). We consider now an iterative algorithm that minimizes G over its domain, whenever $\beta_j + s_j > 0$ and $\beta_j \gamma_j \geq 0$ for each j. Using these conditions, we can rewrite the minimization problem as follows.

Let $d_j = \beta_j + s_j > 0$ and $c > 1$. Let $\hat{P}$ be the matrix with entries $\hat{P}_{nj} = c^{-1} P_{nj}$. Let $\hat{\beta}_j = d_j - c^{-1} s_j$ and $\hat{\gamma}_j$ such that $\hat{\beta}_j \hat{\gamma}_j = \beta_j \gamma_j$. Then it is easily shown that

$$G(x) = \mathrm{KL}(y, \hat{P}x) + \sum_{j=1}^{J} \hat{\beta}_j \mathrm{KL}(\hat{\gamma}_j, x_j) + \text{constants}. \tag{5.1}$$

Consequently, by choosing c large enough, we can make $\hat{\beta}_j > 0$ and $\hat{\gamma}_j \geq 0$ for each j. It is clear then that $G(x)$ attains its minimum value on its domain. This minimization problem is slightly more general than the one arising in the MAP method of [20], in that $\hat{\gamma}_j = 0$ is permitted here. Our iterative algorithm is the following, which is formally the same as in [20]: with $x^0 > 0$ in X, $\hat{s}_j = \sum_{n=1}^{N} \hat{P}_{nj}$ and $k = 0, 1, \ldots$ let

$$x_j^{k+1} = \hat{\alpha}_j \hat{s}_j^{-1} x_j^k \sum_{n=1}^{N} \hat{P}_{nj} \frac{y_n}{(\hat{P}x^k)_n} + (1 - \hat{\alpha}_j)\hat{\gamma}_j \tag{5.2}$$

with $\hat{\alpha}_j = \hat{s}_j / d_j$. In the notation of the original problem, the iterative step becomes

$$x_j^{k+1} = \alpha_j s_j^{-1} x_j^k \sum_{n=1}^{N} P_{nj} \frac{y_n}{(Px^k)_n} + (1 - \alpha_j)\gamma_j \tag{5.3}$$

with $\alpha_j = s_j / d_j$. From (5.2), it is clear that $x_j^k > 0$ for each j and each k. We can also see this from (5.3), since $\beta_j + s_j > 0$ and $\beta_j \gamma_j \geq 0$ for each j.

In the Appendix, we prove convergence of the iterative scheme in (5.3). We then have the following result:

Theorem 5.1: Let the following conditions hold: 1) $\beta_j \gamma_j \geq 0$ for each j and 2) $s_j + \beta_j > 0$ for each j. Then the sequence generated by (5.3) converges to a nonnegative minimizer of the function $G(x)$ given by (4.8).

This theorem was proved in [20], using ideas from [19], for the case in which $N \geq J$, $\beta_j > 0$ and $\gamma_j > 0$ for each j; a simplified proof was given in [4] for the special case of $\beta_j = \beta \geq 0$ for all j. It is important to note that the theorem above permits $\gamma_j = 0$, and so is slightly more general than the result in [20]; we shall need this when we rederive the method of Parra and Barrett as a particular case of our algorithm.

Multiplying by $d_j = s_j + \beta_j$ on both sides of (5.3), we find that

$$\sum_{j=1}^{J} x_j^{k+1} d_j = y_+ + \sum_{j=1}^{J} \beta_j \gamma_j. \tag{5.4}$$

We consider now several particular cases of this algorithm.

VI. Some Particular Cases of the Algorithm

Several well known iterative reconstruction methods are particular cases of the algorithm given in the previous section.

A. The EMML Algorithm for SPECT

The general *expectation maximization* (EM) method for likelihood maximization was presented by Dempster, Laird and Rubin in [13]. Application of this general formalism to the specific case of emission tomography and Poisson data resulted in what, in medical imaging, is usually called the EMML (or, sometimes, the MLEM or just EM) algorithm [29], [30], [19], [20].

For $j = 1, \ldots, J$ and $i = 1, \ldots, I$, let $P_{ij} \geq 0$ be the probability that a photon emitted from the jth voxel is detected at the ith detector and let Y_{ij} be the random number of such detections. The Y_{ij} are assumed to be independent and Poisson distributed, with mean values $P_{ij=x_j}$. The random number of detections made at the ith detector, denoted Y_i, is then Poisson distributed, with mean value $Px_i = (Px)_i = \sum_{j=1}^{J} P_{ij} x_j$. The probability that a photon emitted at the jth voxel will be detected is $d_j = s_j = \sum_{i=1}^{I} P_{ij}$. With $y_i > 0$ the actual

number of detections made at the ith detector, the likelihood function of the vector x becomes

$$L(x) = \prod_{i=1}^{I} \exp(-Px_i)\,(Px_i)^{y_i}/y_i! \tag{6.1}$$

and so maximizing the likelihood, subject to $x_j \geq 0$, is equivalent to minimizing the function $\mathrm{KL}(y, Px)$, subject also to $x_j \geq 0$. Applying our iterative algorithm, with $x_j^0 > 0$ and $\beta_j = 0$ for each j, we have

$$x_j^{k+1} = s_j^{-1} x_j^k \sum_{i=1}^{I} P_{ij} \frac{y_i}{(Px^k)_i} \tag{6.2}$$

this is the EMML algorithm.

It is important to note that although this iterative method was originally derived as a special case of the Dempster, Laird and Rubin algorithm given in [13], it can be viewed apart from any statistical consideration, as an iterative method for finding a nonnegative vector x that minimizes $\mathrm{KL}(y, Px)$, for any positive vector y and any matrix P having nonnegative entries. Accordingly, whenever we encounter an iteration of the form (6.2) we can determine immediately that it is minimizing $\mathrm{KL}(y, Px)$.

In a later section, we shall rederive this algorithm, using only the assumption that the total counts are Poisson distributed, by converting list-mode data to bin-mode.

B. The MAP Method With Prior Gamma Distributions

Lange *et al.* [20] regularize the likelihood maximization problem in emission tomography using a Bayesian approach, in which each of the unknown values x_j is given a prior gamma distribution, with mean μ_j and variance σ_j^2. With $\beta_j = \mu_j/\sigma_j^2$ and $0 < \gamma_j = (\mu_j - \sigma_j^2)/\mu_j$, the likelihood function becomes $G(x)$ as given by (4.8) and their resulting method has the iterative step given in (5.3), with the restriction that $\beta_j > 0$ and $\gamma_j > 0$ for all j.

C. The List-Mode EM Algorithm

As we saw earlier, the list-mode likelihood maximization problem is equivalent to minimizing the function

$$F(x) = \mathrm{KL}(u, Px) + \sum_{j=1}^{J} (d_j - s_j) x_j \tag{6.3}$$

over vectors x with nonnegative entries. Since the functions $f_j(\cdot)$ may be pdf, s_j can be any positive number, so $\beta_j = d_j - s_j$ need not be positive; however, we do have $\beta_j + s_j = d_j > 0$. Applying our iterative method, we obtain the following:

$$x_j^{k+1} = d_j^{-1} x_j^k \sum_{n=1}^{N} P_{nj} \frac{1}{(Px^k)_n}. \tag{6.4}$$

In [26], Parra and Barrett use the original EM formulation of to derive this version of the EMML algorithm applicable to list-mode emission tomographic image reconstruction. The functions $f_j(\cdot)$ in their formulation are pdf. In Huesman *et al.* [16], the same algorithm is derived starting with a bin-mode formulation and the Poisson model and then rewriting the likelihood in list-mode format.

It follows from (5.4) that the list-mode case limit vector x^∞ has the property $x_+^\infty d = \sum_{j=1}^{J} x_j^\infty d_j = y_+ = N$; so, for the solution vector, the expected number of detected events equals the actual number of events on the list.

It is helpful to note that even though the list-mode EM algorithm is derived here as a particular case of an iterative scheme formally identical to a regularization method, that is, the MAP approach of Lange *et al.*, having $\gamma_j = 0$ can mean that an additional regularization term may be needed to achieve a smooth reconstruction. The snowy ML reconstructions one often sees when the EMML algorithm is applied to noisy data can result when the value zero is assigned to many of the voxels in the image, in an apparently random fashion. The images can be improved by regularization methods, such as the MAP method of Lange *et al.*, that prevent zero voxel values; zero values are not prevented in the list-mode EM algorithm, hence noisy reconstructions are still possible.

VII. Bin-Mode Data in List-Mode Form

Suppose now that we have predetermined outcomes, or bins, indexed by $i = 1, \ldots, I$, and that each datum in the list is simply the index number of the bin associated with that item; that is, $z_n = i_n \in \{1, 2, \ldots, I\}$ for each n. Let $y_i > 0$ denote the number of times that the index i occurs in the list. We assume, as before, that the items in the list were generated independently and that N, the number of items in the list, is Poisson distributed, with mean $x_+ d$. We make no further assumptions about the distributions of the counts y_i.

In this case, the set Z of possible events contains the set of bin indexes, but it does not make sense to think of Z as equal to this set. We must also account for events that are not detected. One way to do this is to let Z consist of the indexes $Z = \{1, 2, \ldots, I, I+1\}$, with the event $z = I + 1$ corresponding to an undetected event. Then $\sum_{i=1}^{I+1} f_j(i) = 1$, but $\sum_{i=1}^{I} f_j(i) = s_j \leq 1$ for each j. It makes sense now to let $d_j = s_j$. Since all we have are bin counts y_i, for $i = 1, 2, \ldots, I$, when we convert the bin-mode data to list-mode, the index $z = I + 1$ never appears on the list.

The log likelihood function $LL(x)$ can then be rewritten as

$$LL(x) = -x_+ d + N \log(x_+ d) + \sum_{i=1}^{I} y_i \log(Px)_i \tag{7.1}$$

where P is now the matrix with entries $P_{ij} = f_j(i)$ and $(Px)_i = \sum_{j=1}^{J} P_{ij} x_j$. Then the resulting iterative algorithm is exactly the EMML method. The EMML update equation has been used elsewhere for probabilistic mixtures without the Poisson assumption; see, for example, Lucy [22].

It is also possible to go the other way and to formulate list-mode data in bin-mode form. Typically, this involves loss of information, as distinguishable events are placed in the same bin.

VIII. Block-Iterative Versions of our Algorithm

In recent years, block-iterative or ordered-subset versions of several algorithms have been used in order to accelerate the production of usable images (see, for example, [8] and the refer-

ences given there, as well as [25]). Such methods are based on the decomposition of the index set $\{n = 1, \ldots, N\}$ into (not necessarily disjoint) subsets or blocks, B_t, $t = 1, \ldots, T$. At each step of the iterative procedure we select a current block and we use only those data values y_n with n in the current block. Such methods typically do not converge to a single limit vector, but instead, have subsequential convergence to a limit cycle of (generally) T distinct vectors. Proof of such subsequential convergence has been given only for the simplest algorithms based on least squares distance and no proof of subsequential convergence is known for algorithms based on the KL distance.

For $t = 1, \ldots, T$ and $j = 1, \ldots, J$, let $s_{jt} = \sum_{n \in B_t} P_{nj}$ and let

$$m_t = \max\{s_{jt}/s_j | j = 1, \ldots, J\}.$$

Our block-iterative version of the algorithm in (5.3) is the following.

Algorithm 8.1: Let x^0 be an arbitrary vector in X with positive entries. For $k = 0, 1, \ldots$, let $t(k) = k(\mathrm{mod}\, T) + 1$. If, for $k = 1, 2, \ldots$ we have $x^k \in X$, let the iterative step be given by

$$x_j^{k+1} = a_{jk} x_j^k + b_{jk} s_{jt(k)}^{-1} x_j^k \sum_{n \in B_{t(k)}} P_{nj} \frac{y_n}{(Px^k)_n} + c_j \gamma_j \tag{8.1}$$

where

a_{jk} $(s_j - m_{t(k)}^{-1} s_{jt(k)})/(s_j + \beta_j)$;

b_j $m_{t(k)}^{-1} s_{jt(k)}/(s_j + \beta_j)$;

c_j $\beta_j/(s_j + \beta_j)$.

With $\beta_j = 0$, this algorithm becomes the RBI-EMML method discussed in [7].

As with other block-iterative methods, the design of the blocks and the order in which they are used will affect the degree of acceleration achieved. For a discussion of related issues, see the recent article by Levkovitz *et al.* [21].

IX. Probabilistic Mixtures

We noted with reference to (4.1) that we are using a probabilistic mixture model with unknown mixing coefficients. To illustrate the notion of a probabilistic mixture problem, we present the following simple example.

Suppose there are 11o coincide with the possible outcomes from rolling two dice. The marbles in each bowl come in I different colors, with $p_j(i)$ the known proportion of marbles in the jth bowl having the ith color; note that, for each fixed j, we have $\sum_{i=1}^I p_j(i) = 1$. A pair of (possibly weighted) dice is rolled and one marble is selected from the bowl whose index j is the number shown on the dice. The index i associated with the color of the selected marble is recorded and then the marble is returned to the bowl. This experiment is repeated many times.

Consider now the random variable Z, whose value is the index of the color of the next marble selected. The probability that Z will take the value i, call it $p(i)$, depends both on what the dice show and what the selected bowl contains. This is a mixture. We have

$$p(i) = \sum_{j=2}^{12} w_j p_j(i) \tag{9.1}$$

where w_j is the probability that j will be rolled on the dice. The mixture problem is to estimate these w_j on the basis of observation of the color of the many selected marbles.

Suppose that the experiment is repeated N times and the list of (indexes of) observed colors is $i_1, i_2, \ldots, i_N$. Let w denote the column vector with entries w_j. The likelihood function in this case is then

$$L(w) = \prod_{n=1}^N p(i_n). \tag{9.2}$$

We can rewrite $L(w)$ as

$$L(w) = \prod_{i=1}^I p(i)^{y_i} \tag{9.3}$$

where y_i denotes the number of times the index i appears in the list. Maximizing $L(w)$ is equivalent to maximizing the log likelihood, given by

$$LL(w) = \sum_{i=1}^I y_i \log p(i). \tag{9.4}$$

The problem then is to maximize

$$\sum_{i=1}^I y_i \log \left(\sum_{j=2}^{12} w_j p_j(i) \right)$$

subject to $\sum_{j=2}^{12} w_j = 1$ and $w_j \geq 0$ for all j. We can calculate this maximizer using our iterative algorithm.

We shall consider only those mixture coefficient vectors w in the set

$$W = \left\{ w \left| \sum_{j=1}^J w_j p_j(i) > 0, \qquad i = 1, \ldots, I \right. \right\}.$$

Let $q_i = y_i/y_+$ for each i.

Conveniently, the solution of the problem, to maximize

$$\sum_{i=1}^I \left[q_i \log \left(\sum_{j=1}^J w_j p_j(i) \right) - \sum_{j=1}^J w_j p_j(i) \right]$$

subject only to $w_j \geq 0$, $j = 1, \ldots, J$, has the property that $\sum_{j=1}^J w_j = 1$, since, for $j = 1, \ldots, J$, we have $\sum_{i=1}^I p_j(i) = 1$. Therefore, we can obtain our desired estimate of w by performing this maximization, using the algorithm given by (5.3), with i in place of n, q_i in place of y_n, $\beta_j = 0$ and $P_{ij} = p_j(i)$ in place of P_{nj}. The iterative step becomes

$$w_j^{k+1} = w_j^k \sum_{i=1}^I P_{ij} \frac{q_i}{(Pw^k)_i}. \tag{9.5}$$

In the probabilistic mixture problem described by (4.1), the functions $f_j(\cdot)$ are assumed known and we need only

estimate the mixing coefficients, x_j/x_+. More complicated mixtures are often used in image processing and elsewhere. In [28], Samadani employs a mixture model in which $J = 3$ and the $f_j(\cdot)$ have the known form of gamma distributions, but involve unknown parameters, which must also be estimated. His iterative algorithm has two alternating steps, the first of which updates the mixing coefficients using the same method we presented here.

In [10], a somewhat different approach to such complicated mixture problems was presented. Instead of using a small value of J and functions $f_j(\cdot)$ involving unknown parameters, one can consider a much larger J and a family of functions $f_j(\cdot)$ each associated with a known set of parameter values. Our assumption then is that the mixing coefficients are nonzero for only a few values of j and the objective is to determine which values of j these are. High-resolution spectrum analysis methods were modified to provide estimates of the mixing coefficients that are predisposed to single out only a small number of nonzero values.

X. Closing Remarks

The main goal in this paper has been to provide an optimization-theoretic derivation and proof of convergence of the list-mode EM algorithm. We have shown that the list-mode EM algorithm is a special case of a more general iterative method that also includes the EMML method of SPECT and the MAP *gamma priors* procedure of Lange *et al.* Our general algorithm is formally an extension of this MAP procedure, in which the parameters are allowed to assume values outside the gamma distribution model. The proof of convergence is also more general, and does not require unique minimizers or any restrictions on the values of N and J. Since $\gamma_j = 0$is permitted, solutions on the boundary are once again possible, unlike in the MAP case.

For the list-mode ET problem, the method presented here is essentially the same as that given by Bouwens *et al.* [2], [3]. As they point out, we do not need to know the individual functions $f_j(\cdot)$ or even their values $f_j(z_n)$ at the items in the list; all we need are the values $f_j(z_n)/t_n$, with $t_n = \sum_{m=1}^{J} f_m(z_n)$. The method in [26] does require specification of the densities $f_j(\cdot)$, which then requires detailed description of what information is contained in the entries of each list item. In [26], the authors give an analytic expression for $f_j(z_n)$ for 2-D PET reconstruction, in which each list item contains two 2-D position coordinates and the time-of-flight. The method here, as modified by Bouwens, *et al.*, requires, for each fixed n, only the specification of a finite vector of probabilities describing the relative likelihoods that an emission at any one of the various pixel locations was responsible for the nth list item.

One important aspect of the list-mode problem that we have not touched on here is the actual determination of the various probabilities needed. This issue is treated in some detail in [26], as well as in [2] and [3].

Appendix
Proof of the Theorem

We present here a complete, elementary and self-contained proof of convergence of our iterative algorithm, which, as we have seen, contains the EMML method, the *gamma prior* MAP method and the list-mode EM method as particular cases.

As we saw above, minimizing the function $G(x)$ given by (4.8) is equivalent to minimizing a related function in which the β_j are replaced by positive quantities. From this rewriting, we conclude, first, that the function G does attain its minimum value, and second, that the x^k generated by the algorithm are positive for each k and j. In the proof that follows, we are assuming only that $\beta_j + s_j > 0$ and $\beta_j\gamma_j \geq 0$ for each j.

For any nonnegative vector x in the set $X = \{x|(Px)_n > 0, n = 1, 2, \ldots, N\}$, let $r(x)$ and $q(x)$ denote the N by J arrays with entries $r(x)_{nj} = x_j P_{nj} y_n/(Px)_n$ and $q(x)_{nj} = x_j P_{nj}$, respectively. For any nonnegative vectors x and z in X, define

$$H(x, z) = \mathrm{KL}(r(x), q(z)) + \sum_{j=1}^{J} \beta_j \mathrm{KL}(\gamma_j, x_j). \quad (10.1)$$

For fixed x in X, denote by x' the vector with entries given by

$$x'_j = \alpha_j s_j^{-1} x_j \sum_{n=1}^{N} P_{nj} \frac{y_n}{(Px)_n} + (1 - \alpha_j)\gamma_j. \quad (10.2)$$

From Fact 2 below, we conclude that $z = x'$ is the value of z for which $H(x, z)$ is minimized. It is easy to see that $(x^k)' = x^{k+1}$ as given by (5.2); it is in this sense that we consider the algorithm to be one of sequential unconstrained minimization [15]. Note also that $\mathrm{KL}(r(x), q(x)) = \mathrm{KL}(y, Px)$ and $H(x, x) = G(x)$. The following useful identities are established by tedious, but elementary, calculations, which we omit.

Fact 1:

$$\mathrm{KL}(r(x), q(z)) = \mathrm{KL}(r(z), q(z)) + \mathrm{KL}(r(x), r(z)).$$

Fact 2:

$$H(x, z) = H(x, x') + \sum_{j=1}^{J} (s_j + \beta_j)\mathrm{KL}(x'_j, z_j).$$

From (5.4), we obtain

$$\sum_{j=1}^{J} (s_j + \beta_j) x_j^{k+1} = \sum_{n=1}^{N} y_n + \sum_{j=1}^{J} \beta_j\gamma_j$$

from which it follows that the sequence $\{x^k\}$ is bounded. Let x^* denote an arbitrary cluster point of the sequence $\{x^k\}$.

Using the previous two facts, we obtain

Fact 3:

$$G(x^k) = G(x^{k+1}) + \mathrm{KL}\left(r(x^k), r(x^{k+1})\right) + \sum_{j=1}^{J} (s_j + \beta_j)\mathrm{KL}\left(x_j^{k+1}, x_j^k\right).$$

From Fact 3, it follows that the sequence $\{G(x^k)\}$ is decreasing and the sequence $\{\sum_{j=1}^{J} (s_j + \beta_j)\mathrm{KL}(x_j^{k+1}, x_j^k)\}$ converges to zero; consequently $G(x^*) < +\infty$, so x^* is in X, and x^* is a fixed point of the iteration, that is, $(x^*)' = x^*$.

Now let $\hat{x}$ be a nonnegative minimizer of the function $G(x)$; since $G(\hat{x}) < +\infty$, we know $\hat{x}$ is in X. It follows from Fact

2 that $\hat{x}$ is also a fixed point of the iteration. The proof of convergence is based on using the facts above to write the quantity $H(\hat{x}, x^k)$ in two equivalent ways. Specifically, we have

$$H(\hat{x}, x^k) = G(\hat{x}) + \sum_{j=1}^{J} (s_j + \beta_j)\mathrm{KL}(\hat{x}_j, x_j^k) \quad (10.3)$$

and

$$H(\hat{x}, x^k) = F(x^k) + \mathrm{KL}\left(r(\hat{x}), r(x^k)\right). \quad (10.4)$$

Since $G(x^k) \geq G(\hat{x})$ it follows that

$$\sum_{j=1}^{J} (s_j + \beta_j)\mathrm{KL}(\hat{x}_j, x_j^k) \geq KL\left(r(\hat{x}), r(x^k)\right). \quad (10.5)$$

The key step in the proof is to obtain the following inequality:

$$\mathrm{KL}\left(r(\hat{x}), r(x^k)\right) \geq \sum_{j=1}^{J} (s_j + \beta_j)KL(\hat{x}_j, x_j^{k+1}). \quad (10.6)$$

Combining inequalities (10.5) with (10.6) will produce what we have called elsewhere the *double inequality* on which the proof rests. Inequality (10.6) appears in various guises in other papers; in Csiszár and Tusnády [12] it is their *four-point property*, while in Iusem [17] it is a consequence of the joint convexity of the KL distance. Indeed, it has been shown by Eggermont and LaRiccia that joint convexity of more general Bregman distances implies the four-point property.

The next two lemmas concerning the KL distance are helpful at this stage.

Lemma 10.1: For any nonnegative vectors x and z, with $x_+ = \sum_{j=1}^{J} x_j$ and $z_+ = \sum_{j=1}^{J} z_j > 0$, we have

$$\mathrm{KL}(x, z) = \mathrm{KL}(x_+, z_+) + \mathrm{KL}(x, (x_+/z_+)z).$$

Proof: The proof is an easy calculation. ■

Lemma 10.2: For any positive scalars a, b, and c, with $a \neq b$, we have

$$\mathrm{KL}(a, b) > \mathrm{KL}(a + c, b + c).$$

Proof: With a and b fixed, the function $g(c) = \mathrm{KL}(a + c, b + c)$ has derivative $g'(c) = 1 - t + \log t < 0$, for $t = (a + c)/(b + c)$, so $g(c)$ is strictly decreasing for $c \geq 0$. ■

Applying the first lemma to the sum over the index n, we obtain

$$\mathrm{KL}\left(r(\hat{x}), r(x^k)\right) \geq \sum_{j=1}^{J} \mathrm{KL}\left(\sum_{n=1}^{N} r(\hat{x})_{nj}, \sum_{n=1}^{N} r(x^k)_{nj}\right). \quad (10.7)$$

This can be rewritten as

$$\mathrm{KL}\left(r(\hat{x}), r(x^k)\right) \geq \sum_{j=1}^{J} \mathrm{KL}\left((s_j + \beta_j)\hat{x}_j - \beta_j\gamma_j, \cdot(s_j + \beta_j)x_j^{k+1} - \beta_j\gamma_j\right). \quad (10.8)$$

Applying the second lemma and using the fact that $\beta_j\gamma_j \geq 0$, we get the inequality (10.6).

Combining inequalities (10.5) and (10.6), we conclude that the sequence

$$\left\{\sum_{j=1}^{J} (s_j + \beta_j)\mathrm{KL}(\hat{x}_j, x_j^k)\right\}$$

is decreasing and that $\mathrm{KL}(\hat{x}_j, x_j^*) < +\infty$ for each j. Therefore, $\hat{x}_j > 0$ implies $x_j^* > 0$. Also

$$\sum_{j=1}^{J} (s_j + \beta_j)\mathrm{KL}(\hat{x}_j, x_j^*) = \mathrm{KL}\left(r(\hat{x}), r(x^*)\right) < +\infty.$$

We then have

$$H(\hat{x}, x^*) = G(\hat{x}) + \sum_{j=1}^{J} (s_j + \beta_j)\mathrm{KL}(\hat{x}_j, x_j^*) \quad (10.9)$$

from (10.3) and

$$H(\hat{x}, x^*) = G(x^*) + \mathrm{KL}(r(\hat{x}), r(x^*)) \quad (10.10)$$

from (10.4), from which we conclude that $G(\hat{x}) = G(x^*)$. Therefore, x^* is a nonnegative minimizer of $G(x)$ and we can replace the generic $\hat{x}$ with x^* in the results just obtained. In particular, we now know that the sequence

$$\left\{\sum_{j=1}^{J} (s_j + \beta_j)\mathrm{KL}(x_j^*, x_j^k)\right\}$$

is decreasing; but since a subsequence converges to zero, the entire sequence converges to zero, which tells us that $x^k \to x^*$. This concludes the proof of the theorem. ■

ACKNOWLEDGMENT

The author wishes to thank L. Bouwens, Ghent University, Belgium, who supplied references [2] and [3], and S. Glick and M. King of the University of Massachusetts Medical School for several helpful discussions. The author also thanks an anonymous referee for the idea of rewriting the function $G(x)$ in terms of positive parameters.

The contents of this paper are solely the responsibility of the author and do not necessarily represent the official views of NIH.

REFERENCES

[1] H. Barrett, T. White, and L. Parra, "List-mode likelihood," *J. Opt. Soc. Amer. A*, vol. 14, pp. 2914–2923, 1997.

[2] L. Bouwens, R. Van de Walle, H. Gifford, M. King, I. Lemahieu, and R. Dierckx, "LMIRA: List mode iterative reconstruction algorithm for SPECT," IEEE Transactions in Nuclear Science, Aug. 2001, to be published.

[3] L. Bouwens, H. Gifford, R. Van de Walle, M. King, I. Lemahieu, and R. Dierckx, "Resolution recovery for list-mode reconstruction in SPECT," Physics in Medicine and Biology, Aug. 2001, to be published.

[4] C. Byrne, "Iterative image reconstruction algorithms based on cross-entropy minimization," *IEEE Trans. Image Processing*, vol. IP-2, pp. 96–103, Jan. 1993.

[5] ——, "Erratum and addendum to 'Iterative image reconstruction algorithms based on cross-entropy minimization'," *IEEE Trans. Image Processing*, vol. IP-4, pp. 225–226, 1995.
[6] ——, "Iterative reconstruction algorithms based on cross-entropy minimization," in *Image Models (and their Speech Model Cousins)*. ser. The IMA Volumes in Mathematics and its Applications, S. E. Levinson and L. Shepp, Eds. New York: Springer-Verlag, 1995, vol. 80, pp. 1–11.
[7] ——, "Block-iterative methods for image reconstruction from projections," *IEEE Trans. Image Processing*, vol. IP-5, pp. 792–794, May 1996.
[8] ——, "Block-iterative interior point optimization methods for image reconstruction from limited data," *Inverse Problems*, vol. 16, pp. 1405–1419, 2000.
[9] C. Byrne and Y. Censor, "Proximity function minimization using multiple Bregman projections, with application to split feasibility and Kullback–Leibler distance minimization," Ann. Oper. Res., 2001, to be published.
[10] C. Byrne, D. Haughton, and T. Jiang, "High-resolution inversion of the discrete Poisson and binomial transforms," *Inverse Problems*, vol. 9, pp. 39–56, 1993.
[11] T. Cover, "An algorithm for maximizing expected log investment return," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 369–373, Feb. 1984.
[12] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statistics and Decisions*, no. 1 suppl., pp. 205–237, 1984.
[13] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.*, ser. B, vol. 39, pp. 1–38, 1977.
[14] P. P. B. Eggermont and V. N. LaRiccia, "On EM-like algorithms with EM-like properties for smoothed minimum distance estimation," Dept. Math. Sci., Univ. Delaware, Newark, Tech. Rep., Mar. 1998.
[15] A. Fiacco and G. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, ser. Classics in Mathematics. Philadelphia, PA: SIAM, 1990.
[16] R. Huesman, G. Klein, W. Moses, J. Qi, B. Ruetter, and P. Virador, "List-mode maximum likelihood reconstruction applied to positron emission mammography (PEM) with irregular sampling," *IEEE Trans. Med. Imag.*, vol. 19, pp. 532–537, May 2000.
[17] A. N. Iusem, "A short convergence proof of the EM algorithm for a specific Poisson model," *Revista Brasileira de Probabilidade e Estatistica*, vol. 6, pp. 57–67, 1992.
[18] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
[19] K. Lange and R. Carson, "EM reconstruction for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, pp. 306–312, 1984.
[20] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Trans. Med. Imag.*, vol. MI-6, pp. 106–114, Apr. 1987.
[21] R. Levkovitz, D. Falikman, M. Zibulevsky, A. Ben-Tal, and A. Nemirovski, "The design and implementation of COSEM, an iterative algorithm for fully 3-D listmode data," *IEEE Trans. Med. Imag.*, vol. MI-20, pp. 633–642, June 2001.
[22] L. Lucy, "An iterative technique for the rectification of observed distributions," *Astronom. J.*, vol. 79, pp. 745–754, 1974.
[23] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, ser. Probability and Statistics. New York: Wiley, 1997.
[24] X.-L. Meng and D. van Dyk, "The EM algorithm—An old folk-song sung to a fast new tune," *J. Roy. Statist. Soc. B*, vol. 59, no. 3, pp. 511–567, 1997.
[25] M. Narayanan, C. Byrne, and M. King, "An interior point iterative reconstruction algorithm incorporating upper and lower bounds, with application to SPECT transmission imaging," *IEEE Trans. Med. Imag.*, vol. MI-20, pp. 342–353, Apr. 2001.
[26] L. Parra and H. Barrett, "List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET," *IEEE Trans. Med. Imag.*, vol. 17, pp. 228–235, Apr. 1998.
[27] R. Redner and H. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, pp. 195–239, 1984.
[28] R. Samadani, "A finite mixtures algorithm for finding proportions in SAR images," *IEEE Trans. Image Processing*, pp. 1182–1186, Apr. 1995.
[29] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-1, pp. 113–122, Apr. 1982.
[30] Y. Vardi, L. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Statist. Assoc.*, vol. 80, pp. 8–37, 1985.
[31] C. F. J. Wu, "On convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, pp. 95–103, 1983.