A Generalized EM Algorithm for 3-D Bayesian Reconstruction from Poisson Data Using Gibbs Priors

TOM HEBERT AND RICHARD LEAHY, MEMBER, IEEE

Abstract—A generalized expectation-maximization (GEM) algorithm is developed for Bayesian reconstruction based upon locally correlated Markov random field priors in the form of Gibbs functions and upon the Poisson data model. For the $M$-step of the algorithm, a form of coordinate gradient ascent is derived. The algorithm reduces to the EM maximum likelihood algorithm as the Markov random field prior tends to a uniform distribution. Three different Gibbs function priors are examined. Reconstructions of 3-D images obtained from the Poisson model of single photon emission computed tomography (SPECT) are presented.

I. INTRODUCTION

While a maximum likelihood (ML) approach to emission tomography has attracted considerable interest since the application of the expectation-maximization (EM) formulation by Shepp and Vardi [16], in practice two drawbacks are encountered: 1) due to the ill-conditioned nature of the reconstruction problem, reconstructions tend to take on an increasingly nonsmooth quality as the ML solution is approached; 2) convergence can be increasingly slow as the ML solution is approached. These drawbacks are often compensated for by initializing the algorithm with a smooth estimate and terminating the algorithm before convergence [7], [18]. An alternative regularization approach to the problem based on Grenanders method of sieves was taken by Snyder and Miller in [17] and Miller, Snyder, and Moore in [12]. A MAP approach based on a particular Gibbs prior, which we examine later, was taken by Geman and McClure. In [5] they impose the constraint that the unknown image pixels take values from a known interval and use gradient ascent to arrive at 2-D MAP estimates. In [1], Besag reported that his ICM algorithm had been applied in a preliminary way to gamma camera scans, but no details were provided. The iterated conditional modes (ICM) algorithm [1] is equivalent to coordinate descent [11] of the negative log of the posterior function. This approach updates a single pixel at a time by maximizing a univariate function which is conditioned on all data values to which that pixel contributes, all pixels which contribute to those data values, and all neighbors of the pixel to be updated. In emission tomography, each source pixel contributes to many data values so that this algorithm can be more computationally demanding than the approach presented in this paper. More recently, several authors have investigated Gaussian and Poisson priors applying the EM approach to generate Bayesian reconstruction algorithms. Liang and Hart examined uncorrelated Gaussian and Poisson priors in [6] and in [10, sect. 2] and correlated Gaussian priors in [10, sect. 3]. The algorithms developed in those sections initially follow an EM derivation. However, in order to develop a closed-form $M$-step, they replace a set of parameters in the prior distribution with a set of uncompensated image pixel values. Since this does not result in a true EM algorithm, as noted in [9], the question of whether the posterior distribution is increased at each step and the question of convergence for all possible data sets remains open. Their algorithm was shown to outperform the EM likelihood algorithm in a one-dimensional simulation in [10] and later in 2-D simulations in [6]. Levitan and Herman [9] derive a valid EM algorithm for an uncorrelated Gaussian prior and demonstrate a marked improvement over ML estimates in 2-D simulations. For the mean of the prior distribution, Levitan and Herman used a smoothed filtered-backprojection reconstruction. Since it is generally accepted that the structure of images is one of nonstationary mean and local correlations [14], their prior satisfies the first of these two important image attributes. Markov random fields [14], described by Gibbs distributions, capture the property of local correlation and do not require specification of the mean.

For independent Poisson data with a complete/incomplete data representation, we develop a GEM algorithm for MAP reconstruction based upon locally correlated Markov random field priors in the form of Gibbs functions. For the $M$-step of this GEM MAP algorithm, a form of coordinate gradient ascent is derived. Implementation of this algorithm closely follows that of the EM maximum likelihood (EM ML) algorithm [16]. In addition, as the prior tends towards a uniform distribution, this algorithm reduces to the EM ML algorithm. The reconstructed image pixels are constrained to be nonnegative, but are not restricted to a specified finite interval as in [5] and no approximations as in [6] and [10] are used. Three different

Manuscript received June 20, 1988; revised November 30, 1988. The authors are with the Signal and Image Processing Institute, Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089.
IEEE Log Number 8826125.

0278-0062/89/0600-0194$01.00 © 1989 IEEE
Gibbs function priors are examined. We apply this GEM MAP algorithm to estimate the 3-D image parameters in the Poisson model of single photon emission data.

A Note on Notation: In the following sections we represent vectors with bold-lower-case characters and matrices with bold upper-case characters. Subscripts are used to indicate a particular element of the vector or matrix. Vector superscripts are used to indicate the particular iteration at which the vector has been computed.

II. The Generalized EM Approach to Bayesian Reconstruction

The probability distribution of the image vector \( \lambda \) conditioned on the data vector \( y \) is formulated using Bayes rule

\[
f(\lambda | y) = \frac{f(y | \lambda) f(\lambda)}{f(y)}.
\]

For Bayesian estimation, some a priori information is incorporated in the estimate by specification of the prior \( f(\lambda) \). Maximum a posteriori (MAP) estimation selects \( \lambda \) as the estimate which maximizes \( f(\lambda | y) \) or any monotonic function of \( f(y | \lambda) \). Since \( f(y) \) is a constant for a given data vector, MAP reconstruction requires solution of the problem

\[
\max_{\lambda} B(\lambda | y) = \log f(y | \lambda) + \log f(\lambda) \quad (1)
\]

Consider the case where the prior distribution \( f(\lambda) \) is chosen as uniform over the acceptable parameter space. Then \( f(\lambda) \) is simply a constant and MAP estimation and ML estimation are equivalent. MAP estimation offers a potential for reconstruction improvement when prior distributions, more meaningful than the uniform distribution, can be defined. In image reconstruction, priors are sought which lend increased probability to realizations which feature segmented, slowly changing regions and decreased probability to highly erratic images. Such image models should incorporate local interactions while allowing abrupt changes across edges or region boundaries. Gibbs functions have been demonstrated to be useful in this application [1], [4].

The EM algorithm as presented by Dempster, Laird, and Rubin [3] is a general approach to iterative optimization of likelihood or Bayesian functions when the data can be formulated in a complete/incomplete framework. A complete/incomplete data formulation is applicable when data are missing or when the problem has a more natural formulation in terms of a set of unobserved data. At each iteration, the EM approach requires two steps: an expectation step (E-step) followed by a maximization step (M-step). Often, these two steps can be combined into one.

Let \( x \) be the vector of complete, but unobserved, data and \( y \) the vector of incomplete, but observed, data. In order to apply the EM approach, the relationship between the complete data and the incomplete data must be a many-to-one mapping from \( \{ x \} \) to \( \{ y \} \). That is, given a realization \( \hat{x} \), only one particular realization \( \hat{y} \) has a nonzero probability of having occurred. Given a realization \( \hat{y} \), there is a feasible set \( \{ x \} \), with a nonzero probability of having occurred.

It then follows that due to this many-to-one mapping

\[
f(x | y, \lambda) = \frac{f(x | \lambda) I_x(x)}{f(y | \lambda)} \quad (2)
\]

where \( I_x(x) \) is the indicator function which is equal to 1 if \( x \) results in \( y \) and equal to 0 otherwise. In addition, for any \( \lambda^k \)

\[
E_x \{ \log f(y | \lambda) | y, \lambda^k \} = \int \log f(y | \lambda) f(x | y, \lambda^k) \, dx
\]

\[
= \log f(y | \lambda). \quad (3)
\]

Combining (2) and (3) gives

\[
\log f(y | \lambda) = E_x \{ \log f(x | \lambda) | y, \lambda^k \} - E_x \{ \log f(x | y, \lambda) | y, \lambda^k \} \quad (4)
\]

which is a function of \( \lambda \) and \( y \). It follows that with a complete/incomplete data formulation, \( E_x \{ \log f(x | \lambda) | y, \lambda^k \} \) and \( E_x \{ \log f(x | y, \lambda) | y, \lambda^k \} \) are each functions of \( \lambda^k \) but their difference is not. Substitution of (4) into (1) results in an expression for \( B(\lambda | y) \) given a data vector \( y \)

\[
B(\lambda | y) = Q(\lambda | \lambda^k) - E_x \{ \log f(x | y, \lambda) | y, \lambda^k \} \quad (5)
\]

where

\[
Q(\lambda | \lambda^k) = E_x \{ \log f(x | \lambda) | y, \lambda^k \} + \log f(\lambda).
\]

To clarify how the EM approach works, we first note that, from Jensen’s inequality [15], for any \( \lambda^{k+1} \neq \lambda^k \)

\[
E_x \{ \log f(x | y, \lambda^{k+1}) | y, \lambda^k \} \leq E_x \{ \log f(x | y, \lambda^k) | y, \lambda^k \}
\]

with equality if and only if \( f(x | y, \lambda^{k+1}) = f(x | y, \lambda^k) \) almost everywhere [3], [19]. It follows that a sufficient condition for \( B(\lambda^{k+1} | y) > B(\lambda^k | y) \) is \( Q(\lambda^{k+1} | \lambda^k) > Q(\lambda^k | \lambda^k) \) since the second term on the right-hand side of (5) is guaranteed, from Jensen’s inequality, not to decrease.

Beginning with some initial estimate \( \lambda^0 > 0 \), the EM algorithm for MAP estimation thus consists of the following two steps.

The E-step: Form \( E_x \{ \log f(x | \lambda) | y, \lambda^k \} \).

The M-step: Solve \max_{\lambda} Q(\lambda | \lambda^k)

\[
= E_x \{ \log f(x | \lambda) | y, \lambda^k \} + \log f(\lambda). \quad (7)
\]

If the M-step is carried out to a global maximum of the E-step, the approach is termed an EM algorithm. If the M-step is only carried out to ensure \( Q(\lambda^{k+1} | \lambda^k) \geq Q(\lambda^k | \lambda^k) \), the approach is termed a generalized EM (GEM) algorithm.

We note here that a maximum of the posterior distribution (4) is not obtained by a single M-step since the M-step only involves maximization with respect to a portion
of the posterior distribution with the guarantee that the remaining portion is increased but not maximized. This is also the key to why the EM/GEM approach is not guaranteed to achieve a global maximum even if the $M$-step involves a global maximization [19]. As shown above and in [3], an EM/GEM approach ensures an increase in $B(\lambda | y)$ so that, for $B(\lambda | y)$ bounded from above, convergence to some $B^*$ is assured. Continuity of $Q(\lambda | \lambda^*)$ with respect to both $\lambda$ and $\lambda^*$ is sufficient to ensure that all limit points of the sequence $\{\lambda^k\}$ are stationary points of $B(\lambda | y)$ [19]. In general, if $B(\lambda | y)$ is not unimodal and the set of stationary points contains points which are not local maxima, the EM/GEM approach at best only assures convergence of the sequence $\{\lambda^k\}$ to a stationary value. As Wu states [19], this should not be surprising since in such a case no general optimization algorithms are guaranteed to converge to local maxima. The GEM formulation has the same convergence properties as the EM formulation except for a single exception. This exception, which is provided by Wu [19], occurs when it can be demonstrated that any stationary point which is not a local maximum is additionally not a global maximum of the $E$-step. Since the $M$-step of an EM algorithm requires global maximization of the results from the $E$-step, the EM $M$-step would not arrive at a stationary point which was not a local maxima under the above condition. The difficulty in verifying this condition, were it in fact to hold true for a given problem, makes this condition mainly of theoretical interest. From a practical viewpoint, the question of whether to carry out a global maximization versus an increase within the $M$-step is solely one concerned with increasing the per-iteration convergence speed at an increased per-iteration computational cost. For a complete treatment of the EM/GEM approach and its convergence properties, see [3] and [19].

Optimization of likelihoods or Bayesian functions with independent priors may result in closed forms for the $M$-step [9]. If the complete data $x$ are independent, the complete data are a linear function of the incomplete data, and the image pixels $\lambda$ are treated as independent, the $M$-step only requires optimization of a set of univariate functions. However, it is generally accepted that the structure of images is one of local correlation [14]. It would therefore seem more desirable to examine the use of locally correlated priors. When used with an EM approach, correlated priors prohibit the existence of closed-form solutions for an EM $M$-step. Each EM $M$-step thus requires an iterative optimization of an $N$-dimensional function, $N$ being the dimension of $\lambda$. What results is an iterative optimization algorithm within each iteration of an iterative optimization algorithm. In this case, iterative maximization of the posterior distribution by a method such as ICM [1] without using an EM formulation would seem more sensible than an EM approach. The GEM algorithm presented in this paper offers an attractive alternative to both of these approaches.

It should be noted that the use of a nonuniform prior distribution can induce local minima as well as local maxima in both the posterior distribution and in the results from an $E$-step of an EM formulation. Therefore, setting the gradient of the $E$-step equal to zero and solving the resulting set of equations does not ensure a valid $M$-step and may result in jumps to estimates of drastically lower posterior function value. Alternatively, using an approximation to achieve a closed-form $M$-step opens the question of convergence and the possibility at any stage of decreasing the posterior function value. We prefer a generalized EM approach since it guarantees a monotonic increase of the posterior function and has proven convergence properties. Under this approach, each $M$-step may consist of 1 or more iterations of an algorithm to increase $Q(\lambda | \lambda^*)$ without the requirement of maximizing it. We would expect the per-iteration speed of convergence to be slower for a GEM versus an EM approach. However, where a closed-form $M$-step does not exist, the GEM approach may generate estimates with greater increases in the posterior distribution function for a given amount of computation. This may often be the case since it is generally true that the first iteration of an iterative optimization algorithm produces the largest improvement. In this work, we perform Bayesian reconstruction using Markov random field priors in the form of Gibbs functions.

III. NEIGHBORHOOD SYSTEMS AND GIBBS FUNCTION PRIORS

A discrete Markov random field (MRF) defined on a lattice is a collection of random variables, corresponding to the sites of the lattice, for which the probability of a given site value conditioned on the values of all other sites in the lattice is equal to the probability of the site value conditioned on the values at a small subset of the lattice sites. This subset of the lattice sites is called the neighborhood of the given site. Let the set of indexes of sites in the neighborhood of pixel $j$ be denoted $N_j$. Then

$$P(\lambda_i | \lambda_j: i \neq j) = P(\lambda_j | \lambda_i: i \in N_j).$$

If $\lambda_i$ is a neighbor of $\lambda_j$, then $\lambda_j$ is required to be a neighbor of $\lambda_i$. Neighborhoods are referred to as zeroth-order, first-order, $\cdots$, $N$th-order. Fig. 1 shows the zeroth-order, first-order, and second-order neighborhoods for the 2-D lattice and the first-order neighborhood for the 3-D lattice. By the Hammersly–Clifford theorem (1971) [2] a random field defined on a lattice is a Markov random field if and only if its distribution function corresponds to a Gibbs function. To define the form of a Gibbs function, we must first define a clique. A clique is either a single site or a set of sites such that each site in the clique is a neighbor of all other sites in the set. The clique types associated with each neighborhood are also shown in Fig. 1. For a rectangular 2-D lattice and the first-order neighborhood shown in Fig. 1, the cliques are sets of sites consisting of a single site or two horizontally or vertically adjacent sites.

A Gibbs distribution is a probability measure on the set of configurations $\{\lambda\}$ which has the form

$$f(\lambda) = \frac{1}{K} e^{-U(\lambda)/\beta}.$$
where $\beta$ is a constant, $K$ is the normalizing constant (partition function), and $U(\lambda)$ is termed the energy function. The energy function has the form

$$U(\lambda) = \sum_{c \in C} V_c(\lambda)$$

where $C$ denotes the set of all cliques and $V_c(\lambda)$, termed a potential function, is a function of the intensities at the sites contained in clique $c$.

Gibbs functions provide a powerful class of correlated priors. The appeal of the Gibbs prior is that it can be defined to within a normalizing constant by defining a suitable pixel neighborhood and potential functions on the cliques associated with that neighborhood. The intent is to capture the desired properties of the unknown image by a suitable choice of energy function. For Bayesian reconstruction, the normalizing constant, which depends on $\beta$, need not be calculated. The parameter $\beta$ controls the degree to which the modes of the Gibbs prior are accenteduated. As $\beta \to +\infty$ the Gibbs distribution tends to the uniform distribution. As $\beta \to 0$ the prior becomes increasingly more pronounced about its modes. With a Gibbs prior, (1) becomes

$$\max_{\lambda} B(\lambda \mid y) = \log f(y \mid \lambda) - \frac{1}{\beta} \sum_{c \in C} V_c(\lambda)$$

so that as $\beta \to +\infty$ Bayesian reconstruction is unaffected by the prior and reduces to maximum likelihood reconstruction. In this work and in [4] and [5], acceptable values for $\beta$ were obtained by trial and error. Further, as shown in the result section, $\beta$ values from the interval $+\infty$ to some lower limit value can produce an improvement in the reconstruction. Smaller $\beta$ values produce a degradation due to an over-influence by the prior. There is a need for statistical methods of determining optimal $\beta$ values given suitably normalized potential functions and the data set.

Pixel configurations of lowest energy are of highest probability. It is common to choose energy functions which penalize configurations with neighboring pixels differing by large amounts. The Gaussian prior with a diagonal covariance matrix $H$ and mean image $m$ chosen by Levitan and Herman in [8] is also a particular case of a Gibbs prior with a zeroth-order neighborhood. A zeroth-order neighborhood has only cliques containing a single pixel. The corresponding energy function is $-\gamma/2 (\lambda - m)^T H (\lambda - m)$. For the work in this paper, we have chosen a first-order neighborhood (Fig. 1). In [1], a first-order neighborhood is considered unrealistic for most applications. However, in three dimensions a first-order neighborhood results in six neighbors for every pixel versus four in two dimensions. As results presented here indicate, a first-order neighborhood in three dimensions may be sufficient for many applications. In this paper we examine three potential functions. None of the resulting Gibbs priors impose a mean on the image since the potential function on cliques containing a single site have been defined equal to zero.

1) $V_1(\lambda; \lambda_0) = (\lambda - \lambda_0)^2$

2) $V_2(\lambda; \lambda_0) = \frac{(\lambda - \lambda_0)^2}{\delta^2 + (\lambda - \lambda_0)^2}$

3) $V_3(\lambda; \lambda_0) = \log \left(1 + \left(\frac{\lambda - \lambda_0}{\mu}\right)^2\right)$

The first is used by Geman and Geman in [4]. The second is equivalent to that used by Geman and McClure in [5]. The third potential function is a compromise between the first two. A plot of these three potential functions, suitably normalized, versus the difference between the two pixels in a clique is shown in Fig. 2. This normalization simplifies the interpretation of the effects of the priors and enables some comparison for a given $\beta$ value. The first potential function increasingly penalizes the separation between neighboring pixels. In addition, it does so at an increasing rate as the separation increases. To improve on this, we seek a potential function which penalizes separations within uniform regions without unduly penalizing the larger separations which we foresee occurring at the boundary between two different regions of the image. The second prior encourages neighboring pixels to be of similar value until they have become separated through iterative reconstruction by a given threshold $\delta$. At this point the prior allows further separation of their values at a relatively small increase in the penalty. We found that in...
MAP reconstructions of single photon emission images from low total count numbers, some single pixels may separate in value from all their neighboring pixels. When this occurs, the second prior allows this separation to be exaggerated. We found that the quality of the reconstruction could benefit from a compromise between the first and second priors. The third prior represents a compromise which alleviates this problem by increasingly penalizing pixel separation over the full range of separation values.

IV. APPLICATION TO EMISSION TOMOGRAPHY

Reconstructing the distribution of a radioactively tagged compound in a patient is important in assessing organ function and in assessing the efficacy of chemotherapeutic agents. Upon introduction into a patient, a radiopharmaceutical either rapidly or gradually accumulates in various regions of the body. As the nuclei of the radioactive atoms in the compound decay, $\gamma$-rays are emitted. Some of these $\gamma$-rays exit the body and are recorded by the imaging system. A $\gamma$-ray imaging system records $\gamma$-rays which are detected within several hundred thousand separate bins. The number of counts in each bin comprise the data. Corresponding to each bin is a relatively small 3-D volume from which, with high probability, the emission originated. However, these 3-D volumes overlap one another so that any point in the source space may contribute to 50 or more bins. Typically, several million $\gamma$-rays will be detected and registered into several hundred thousand bins. The number of counts in these bins can be shown to be conditionally independent Poisson random variables [16]. To determine the distribution of a radiopharmaceutical, we divide the 3-D space viewed by the imaging system into small volumes called pixels. By reconstructing the Poisson mean of each source pixel, we characterize the distribution of isotopically tagged compound in the patient.

An emission imaging system has a complete/incomplete data formulation in terms of the observed but incomplete data $y$ and complete data set $\{x_{ij}\}$ where $y_i$ represents the number of counts in bin $i$ and $x_{ij}$ represents the number of $\gamma$-ray emissions from source pixel $j$ detected at bin $i$ [16]. Clearly, $y_i = \sum_j x_{ij}$. The complete data $x_{ij}$ are well modeled as independent Poisson random variables with means $P_{ij} \lambda_j$, where $\lambda_j$ is the mean of the total number of $\gamma$-ray emissions from source pixel $j$ and $P_{ij}$ is the probability that a $\gamma$-ray emitted from source pixel $j$ will be detected at camera bin $i$. Thus,

$$f(x|\lambda) = \prod_i \prod_j e^{-\left(P_{ij} \lambda_j\right)} \left(P_{ij} \lambda_j \right)^{y_i} \left(x_{ij}\right)! / \left(x_{ij}\right)!$$

The logarithm of $f(x|\lambda)$ is a linear function of the $x_{ij}$'s plus terms independent of $\lambda$. Therefore, for the $E$-step (6) we only need to compute $E\{x_{ij} | y \lambda^k\}$. The variables $\{x_{iq} | q = 1, \ldots, N\}$ are independent and Poisson with means $P_{iq} \lambda_q$. In addition, $\sum_q x_{iq} = y_i$, so that

$$f \left( x_{i1}, \ldots, x_{iN} | y; \lambda^k \right) = f \left( x_{i1}, \ldots, x_{iN} | y_i; \lambda^k \right).$$

The joint distribution of a set of independent Poisson variables $z$ conditioned on their sum is a multinomial distribution with probability in each class $j$ equal to $E \{z_j\}/\Sigma E \{z_j\}$ [13]. The expected value of a multinomial variable is equal to the probability in its class times the total number of trials. The joint distribution of $\{x_{iq} | q = 1, \ldots, N\}$ conditioned on $\sum_i x_{iq} = y_i$ is

$$f \left( x_{i1}, \ldots, x_{iN} | y_i; \lambda^k \right) = \frac{\left(y_i\right)!}{\left(x_{i1}\right)! \cdots \left(x_{iN}\right)!} \prod_j \left(\mu_{ij}\right)^{x_{ij}} / \prod_j \left(\mu_{ij}\right)^{y_j}$$

where

$$\mu_{ij} = \frac{P_{ij} \lambda_j}{\sum_q P_{iq} \lambda_q}.$$ 

Therefore, $E \{x_{ij} | y \lambda^k\}$, which is the mean of class $j$ of the multinomial distribution, is equal to

$$E_j \{x_{ij} | y \lambda^k\} = \frac{y_j \lambda_j^k}{\sum_q P_{iq} \lambda_q^k}.$$ 

The $E$-step can now be formed as

$$E_k \{ \log f(x|\lambda) | y \lambda^k\} = \sum_j (-a_j \lambda_j + b_j^k \log \lambda_j)$$

$$+ \text{ terms independent of } \lambda^k$$

where

$$a_j = \sum_i P_{ij}$$

and

$$b_j^k = \frac{y_j \lambda_j^k}{\sum_q P_{iq} \lambda_q^k}.$$ 

If the $P_{ij}$'s have been normalized as in [16], then $a_j = 1$ for all $j$. In forming $Q(\lambda^k | \lambda^{k+1})$ of (4) we can omit the terms independent of $\lambda$ since these do not affect the $M$-step (7).

For the generalized EM approach, the $M$-step is to find a
such that $Q(\lambda^{k+1} | \lambda^k) > Q(\lambda^k | \lambda^k)$ where, from (4) and (8),

$$Q(\lambda | \lambda^k) = \sum_j \left(-a_j \lambda_j + b_j \log \lambda_j - \sum_{i \in C} \frac{V_i(\lambda)}{\beta}\right).$$ (9)

The algorithm we present here is a generalized expectation-maximization (GEM) algorithm. It is formulated for Bayesian reconstructions based on the class of correlated priors represented by Gibbs functions. This algorithm is therefore referred to as a GEM MAP algorithm. Its implementation closely follows that of the EM ML algorithm. We note that this algorithm updates pixels sequentially and that updated values are used to update the pixels that follow. Therefore, in Step 2a below, the superscript has been omitted from $\lambda_i$ in $V(\lambda_i; \lambda^j)$ because the neighboring pixels $\lambda_i$ may consist of both updated and un-updated pixels. Step 1 contains the $E$-step of this GEM MAP algorithm and Step 2 contains the $M$-step. Step 2 can be repeated so that each pixel is visited and updated more than once during a GEM MAP iteration. Our experience is that Step 2 need only be performed once or twice after each completed Step 1 to achieve an expeditious rate of convergence. At the $k$th iteration, perform the following steps.

**Step 1**
For all image pixels, compute the usual EM maximum likelihood algorithm [16] updated variables $\lambda^{EM}_j = \frac{b_j}{a_j}$ where $a_j$ and $b_j$ are defined in (8).

**Step 2**
To update the image, visit pixel sites sequentially. When visiting a pixel $\lambda_j$, do Steps 2a–2d.

**Step 2a**
Compute $C_1$ and $C_2$ where

$$C_1 = a_j (\lambda_j^{EM} - \frac{C_2 \lambda_j^k}{a_j})$$

$$C_2 = \frac{1}{\beta} \sum_{i \in N_j} \frac{\partial}{\partial \lambda_j} V(\lambda_i; \lambda_j).$$

**Step 2b**
Set $\alpha = 1$.

Compute $\lambda_j^{k+1} = \lambda_j^{EM} - \frac{C_2 \lambda_j^k}{a_j}$

If $\lambda_j^{k+1} > 0$ go to 2d

If $\lambda_j^{k+1} \leq 0$ compute $\alpha = \frac{0.5}{1 - \frac{\lambda_j^{EM}}{\lambda_j^k} + \frac{C_2 \lambda_j^k}{a_j}}$

**Step 2c**
Compute $\lambda_j^{k+1} = (1 - \alpha) \lambda_j^k + \alpha \left\{ \lambda_j^{EM} - \frac{C_2 \lambda_j^k}{a_j} \right\}$.

**Step 2d**
Check if $a_j (\lambda_j^{k+1} + \lambda_j^{EM} \log \lambda_j^{k+1}) - \sum_{i \in N_j} \frac{V(\lambda_i; \lambda_j^{k+1})}{\beta} \geq C_1$.

If yes, update pixel $j$ to $\lambda_j^{k+1}$ and visit the next pixel. If no, divide $\alpha$ by 2 and return to Step 2c.

To further clarify the steps, when visiting a pixel $j$, Step 2a computes two values which are functions of the data and of the updated and un-updated pixels in the neighborhood of pixel $j$. In Step 2b, if the step-size $\alpha = 1$ results in a negative $\lambda_j^{k+1}$, $\alpha$ is replaced by the step size resulting in $\lambda_j^{k+1} = 0.5 \lambda_j^k$. Step 2c implements a coordinate ascent [11] step for the specified step-size, and Step 2d ensures that the step-size $\alpha$ has resulted in an increase in $Q(\lambda | \lambda^k)$. If $Q(\lambda | \lambda^k)$ has not been increased, Step 2d cuts the step-size in half. As $\beta \to +\infty$, $C_1 \to 0$ and the pixels are updated by setting them equal to the EM ML updated pixel values. Step 2d, which would then represent a check to ensure that the likelihood function has been increased, is always satisfied so that Steps 2a, 2c, and 2d are no longer necessary. As $\beta \to +\infty$, this algorithm thus reduces to the EM ML algorithm.

Let us consider why this is a GEM algorithm. If Step 2d ensures a monotonic increase of $Q(\lambda | \lambda^k)$ and Step 2c is guaranteed to arrive at a $\lambda_j^{k+1}$ satisfying Step 2d, then at the conclusion of the $M$-step $Q(\lambda^{k+1} | \lambda^k) > Q(\lambda^k | \lambda^k)$ and the algorithm is a GEM algorithm. Let the potential function on cliques containing a single pixel be set to zero. For a first-order neighborhood, specify a potential function $V(\lambda_i; \lambda_j)$ evaluated on all cliques containing two pixels. The energy function has the form $1/\beta \sum_{i \in C} V(\lambda_i; \lambda_j)$. Let $N_j$ denote the set of indexes of pixels which are neighbors of pixel $j$. In order to show that (9) is increased at each stage, there are only two cases we must consider: updating a pixel whose neighbors have not been updated in the present iteration, and updating a pixel for which one or more of the neighbors have been updated. Without loss of generality, let us examine Steps 2a–2d for two pixels $\lambda_i$ and $\lambda_j$ which are neighbors by writing $Q(\lambda | \lambda^k)$ explicitly in terms of $\lambda_i$ and $\lambda_j$.

$$Q(\lambda_i; \lambda_j; i \neq j, \lambda_i^k) = -a_i \lambda_i + b_i \log \lambda_i - \sum_{l \in N_j \setminus \{i\}} \frac{V(\lambda_i; \lambda_l)}{\beta} - \frac{V(\lambda_i; \lambda_j)}{\beta}$$

$$= -a_i \lambda_i + b_i \log \lambda_i - \sum_{l \in N_j \setminus \{i\}} \frac{V(\lambda_i; \lambda_l)}{\beta} + \sum_{l \in N_j \setminus \{i\}} (-a_i \lambda_i + b_i \log \lambda_i)$$

$$= -a_i \lambda_i + b_i \log \lambda_i - \sum_{p \neq q} \frac{V(\lambda_p; \lambda_q)}{\beta}.$$ (10)
Let us first visit pixel $i$ whose neighbors have not been updated. Step 2d ensures an increase in the sum of terms 1, 2, 3, and 4 in (10), with no effect on the other terms. Therefore, $\lambda^{k+1}_i$ satisfies

$$Q(\lambda^{k+1}_i; \lambda^i; l \neq i, j | \lambda^k) > Q(\lambda^i; \lambda^i; \lambda^i; l \neq i, j | \lambda^k).$$

For the second case, let us then visit and update pixel $j$, a neighbor of pixel $i$. Step 2d further guarantees an increase in the sum of terms 4, 5, 6, and 7 in (10), with no effect on the other terms. Therefore, $\lambda^{k+1}_j$ satisfies

$$Q(\lambda^{k+1}_j; \lambda^i; \lambda^i; l \neq i, j | \lambda^k) > Q(\lambda^i; \lambda^i; \lambda^i; l \neq i, j | \lambda^k).$$

(11)

Each pixel that is updated according to Step 2d results in an increase in $Q(\lambda | \lambda^k)$ and from Section II, $B(\lambda | y)$ is increased.

It remains to show that Step 2c will result in an updated pixel satisfying Step 2d. Step 2c is a coordinate gradient ascent [11] of $Q(\lambda | \lambda^k)$ initialized at $\lambda^k$. At each step of a coordinate ascent algorithm, only changes to a single element $\lambda_i$ are allowed such that a monotonic increase in $Q(\lambda | \lambda^k)$ is achieved. Each element is addressed in the step order so that for bounded functions ultimate convergence to a stationary value is assured. Let us take the case above, where $\lambda^k_i$ has been updated to $\lambda^{k+1}_i$ and we wish to update $\lambda^k_j$. A coordinate gradient ascent of $Q(\lambda^{k+1}_i; \lambda^i; \lambda^i; l \neq i, j | \lambda^k)$ takes the form

$$\lambda^{k+1}_j = \lambda^k_j + m_j \frac{\partial}{\partial \lambda^j} Q(\lambda^{k+1}_i; \lambda^i; \lambda^i; l \neq i, j | \lambda^k)$$

(12)

where $m_j$ is any positive value. From (10)

$$\frac{\partial}{\partial \lambda^j} Q(\lambda^{k+1}_i; \lambda^i; \lambda^i; l \neq i, j | \lambda^k) = -a_j + \frac{b_j}{\lambda^k_j} - C_z = a_j \left(-1 + \frac{\lambda^EM_j}{\lambda^k_j} - \frac{C_z}{a_j}\right)$$

(13)

where $C_z$ is defined in Step 2a. Let $m_j = a \lambda^k_j / a_j$. Since $\lambda^k$ is constrained to be positive, $0 < a \leq 1$, and $a_j > 0$, $m_j$ is positive. Substituting $m_j$ and (13) into (12) gives Step 2c. The step direction follows the $j$th coordinate directional derivative of $Q(\lambda^{k+1}_i; \lambda^i; \lambda^i; l \neq i, j | \lambda^k)$ and the initial step-size is chosen to mimic the EM likelihood algorithm. An increase in $Q(\lambda^{k+1}_i; \lambda^i; \lambda^i; l \neq i, j | \lambda^k)$ is possible with a sufficiently small step size if that derivative is nonzero. If the derivative is zero, Step 2d is satisfied immediately. Steps sizes which do not satisfy Step 2d are halved so that the algorithm quickly arrives at a step size increasing $Q(\lambda | \lambda^k)$.

This GEM algorithm monotonically increases $Q(\lambda | \lambda^k)$ and terminates at a point $\lambda^k$ for which

$$[\nabla Q(\lambda^k | \lambda^k)] \rightarrow \begin{cases} 0 & \text{if } \lambda^k_j > 0 \\ <0 & \text{if } \lambda^k_j = 0 \end{cases}$$

(14)

for all $j$; i.e., at a point where the directional derivative in all feasible directions is less than or equal to zero. From (4)

$$\nabla B(\lambda | y) = \nabla Q(\lambda | \lambda^k) - \nabla E_z \{ \log f(x | y \lambda) y \lambda^k \}$$

Since $\lambda^k$ maximizes $E_z \{ \log f(x | y \lambda^k) y \lambda^k \}$ (5),

$$\nabla E_z \{ \log f(x | y \lambda^k) y \lambda^k \} = 0$$

(19). This holds regardless of whether or not $\lambda^k$ lies on a boundary. Therefore, $\nabla B(\lambda^k | y) = \nabla Q(\lambda^k | \lambda^k)$ and the algorithm terminates at a point $\lambda^k$ for which the directional derivative of $B(\lambda^k | y)$ in any feasible direction is less than or equal to zero, i.e., $\{\lambda^k\}$ converges to a stationary point.

V. RESULTS

The simulations we present model a 3-D single photon emission imaging system consisting of a parallel collimated gamma camera with $48^3$ pixels viewing a 3-D volume with data collected from 48 different equispaced angles. Perfect collimation is assumed. A 3-D source space consisting of $48^3$ pixels is reconstructed. The Gibbs distribution potential functions were defined as in Section III for a first-order neighborhood consisting of the six nearest neighbors. The neighbors of an interior image pixel consist of the pixels above, below, and on all four sides totaling six neighbors. The missing neighbors of pixels located on the side boundaries are assumed zero. A free boundary [4] is used for pixels on the top and bottom planes of the 3-D reconstruction space. These pixels have fewer neighbors. Pixels on the bottom layer of the 3-D reconstruction space have no neighbor below while pixels on the top layer have no neighbor above.

Fig. 3 shows transverse planes of the 3-D computer generated phantom used in our simulation study. A total mean of 2 million counts was generated from this 3-D phantom. Figs. 4–7 provide a visual comparison of the reconstructions from the EM ML and GEM MAP algorithms with each of the three priors discussed in Section III. The reconstructions displayed in these figures are the result of 50 iterations of each specified algorithm. With more iterations, the EM ML reconstruction deteriorated further while the GEM MAP reconstructions had for all purposes converged. For these reconstructions $\beta = 1$ and the three potential functions were normalized as shown in Fig. 2. Fig. 4, the EM ML reconstruction, shows the excessive nonsmoothness reported by many authors [5], [7], [9], [17], [18]. Figs. 5–7 show that all three Gibbs priors produced a considerable visible improvement in the reconstruction.

Fig. 8 shows the $L_2$ error between the true source image and the reconstruction for 100 iterations of the EM ML algorithm and for 100 iterations of the GEM MAP algorithm using the three different Gibbs priors in Section III. The GEM MAP algorithm initially reduced the $L_2$ error.
as quickly as the EM ML algorithm. The EM ML algorithm characteristically iterated away from the true source image after a number of iterations while the GEM MAP algorithm continued to reduce the $L_2$ error until convergence.

Fig. 9 shows the log of the Bayes function (likelihood) value at each iteration. This shows a monotonic increase in these functions for both the EM ML algorithm and the GEM MAP algorithm. Fig. 10 examines a range of $\beta$ values. Here, the $L_2$ error between the true source image and the reconstruction for 100 iterations of the GEM MAP algorithm using the first potential function $V_1$ with different values of $\beta$ is shown. According to this criterion, we found an improvement in the reconstruction for all values of $\beta \geq 1.0$. 
Gibbs functions and has the desirable theoretical convergence of its generalized EM formulation. For 3-D images, such as those encountered in emission tomography, this algorithm can provide an improvement over maximum likelihood reconstruction at a nominal increase in computational cost. This work shows that some improvement can be achieved for a wide range of $\beta$ values, but statistical methods for optimizing the choice of this parameter are still needed.

REFERENCES


VI. CONCLUSION

The GEM MAP algorithm we have presented can be used with any locally correlated priors in the form of

![Plot of the likelihood and Bayesian functions for 100 EM ML and GEM MAP iterations with $\beta = 1.0$.]

![Plot of the $L_2$ difference between the true 3-D image and the reconstruction for 100 GEM MAP iterations with $V_1$ and different $\beta$ values.]

These simulations were run in Fortran code on a Sun 3/110 workstation with a floating point accelerator. Some improvement in the speed was achieved by setting the parameter space to $\lambda_i \geq \phi > 0$ where $\phi$ was chosen as some small value such as $\phi = 0.0001$. If $\lambda_i$ equaled $\phi$ and the directional derivative was negative along that coordinate, $\lambda_i$ was left equal to $\phi$ and the next pixel was visited. The reduced some of the time spent computing small steps for small pixel values converging towards zero, a task particularly time consuming as the algorithm converges. The EM ML algorithm required 48 s per 3-D iteration while the GEM MAP algorithm averaged 63 s per 3-D iteration. The 3-D forward/back projections required the majority of the CPU time (47 s per iteration) while pixel updating filled the remaining seconds.