Maximum-Likelihood Reconstruction of Transmission Images in Emission Computed Tomography via the EM Algorithm

John M. Ollinger, *Member, IEEE*

Abstract—The expectation-maximization (EM) algorithm for computing maximum-likelihood estimates of transmission images in positron-emission tomography (PET) [1] is extended to include measurement error, accidental coincidences and Compton scatter. A method for accomplishing the maximization step using one step of Newton's method is proposed. The algorithm is regularized with the method of sieves. Evaluations using both Monte Carlo simulations and phantom studies on the Siemens 953B scanner suggest that the algorithm yields unbiased images with significantly lower variances than filtered-backprojection when the images are reconstructed to the intrinsic resolution. Large features in the images converge in under 200 iterations, while the smallest features required up to 2,000 iterations. All but the smallest features in typical transmission scans converge in approximately 250 iterations. The initial implementation of the algorithm requires 50 sec per iteration on a DECStation 5000.

I. INTRODUCTION

ALTHOUGH EMISSION COMPUTED TOMOGRAPHY (ECT) is primarily concerned with emission images, transmission imaging is also of interest. In positron-emission tomography (PET), transmission images are useful for image registration [2] and attenuation correction [3]–[5]. These images are usually reconstructed using filtered-backprojection [6], a method which is less than ideal for emission tomography.

Filtered-backprojection is based on the assumption that the observed data are line integrals through a function whose range is the linear attenuation coefficient. In x-ray computed tomography these data are collected with gas-filled radiation detectors, which measure the number of ionized particles generated by interactions of incident x-rays in a pressurized gas. This yields a signal proportional to both the number of photons and the average absorbed energy, which, along with the relatively high flux of incident photons, suggests the model of a signal in additive Gaussian noise. This Gaussian model is inappropriate for transmission imaging in ECT because the detectors count individual photons at relatively low photon fluences. Photon survival probabilities can be as low as 0.008 in cardiac scans of large individuals, which leads to measurements with fewer than ten counts per projection bin. Hence, a Poisson model is appropriate.

These low counting statistics are a greater problem in transmission imaging than in emission imaging because the transmission data undergo a nonlinear transformation before reconstruction, which introduces singularities and a systematic bias. These errors come about as follows. Two scans are required to compute a transmission image: a blank scan yielding a blank sinogram $B_{\theta k}$, performed with the scanner empty, and a transmission scan yielding a transmission sinogram $T_{\theta k}$, performed with the patient in the scanner. This transmission sinogram is assumed to be corrupted by additive noise, $A_{\theta k}$, arising from accidental coincidences or Compton scatter. The line integrals to be reconstructed, $\hat{s}_{\theta k}$, are given by

$$\hat{s}_{\theta k} = -\ln \frac{T_{\theta k} - A_{\theta k}}{B_{\theta k}} \approx \int_{\ell(\theta k)} \mu(\underline{x})d\ell \qquad (1)$$

where θ is the projection angle, k is the transverse position, and $\ell(\theta k)$ is the line at angle θ and distance k from the origin. The blank scan is assumed to have been corrected for accidental coincidences, and, since it is typically long, can be assumed to be noiseless. The transmission scan, however, must be kept short to minimize patient motion and discomfort, and is often quite noisy. Since the logarithm is only defined for positive numbers, $\hat{s}_{\theta k}$ does not exist for zero or negative values of the numerator of (1). Moreover, for low-count data the logarithm skews the distribution of $\hat{s}_{\theta k}$ such that the estimated line integrals are biased. For noisy transmission data, therefore, the estimated line integrals $\hat{s}_{\theta k}$ are undefined for some projections and biased for others.

Lange and Carson [1] first proposed computing the maximum-likelihood estimate of transmission images via an expectation maximization (EM) algorithm [7] based on Poisson statistics. This approach has merit for the following reasons: it is based on a Poisson model which accurately models low count data; it is constrained to be nonnegative; the logarithm in (1), and hence the singularity, is eliminated; and EM algorithms are frequently easy (albeit slow) to compute. Lange et al. [8] subsequently proposed a simple gradient algorithm for computing the maximum likelihood estimate. This method suffers from the need to perform a line-search at each iteration, which requires costly evaluations of the likelihood function. Browne and Holmes [9] proposed an EM algorithm similar to that in [1] except that different approximations are made in the M-step. Sauer and Bouman [10], [11] proposed an algorithm similar to Gauss-Seidel for finding a maximum *a posteriori* estimate based on a

Markov random field model. Fessler [12] proposed a penalized weighted least square method based on an iterative coordinate descent algorithm. Although least-squares methods are significantly faster than the maximum likelihood algorithms, the quality of the least-squares images relative to the maximum likelihood images is as yet unclear. Mumcuoglu [13] proposed a penalized maximum likelihood algorithm that incorporates Compton scatter and is optimized with a gradient ascent method.

We have extended the EM algorithm proposed by Lange and Carson [1] by incorporating accidental coincidences and measurement error into it and by regularizing it with the method of sieves [14, 15]. Moreover, we make none of the approximations present in the M-steps of the algorithms proposed in [1] and [9]. In this paper, we present these extensions and an evaluation of the algorithm.

II. MATHEMATICAL DEVELOPMENT

Direct maximization of the likelihood function in transmission tomography is difficult because it requires the maximization of a nonlinear function over a large number of variables. The EM algorithm [7] avoids this difficult maximization as follows. The observed data are assumed to lie in an *incomplete data space* with the probability density $g(y|\Phi)$, where Φ is the parameter vector to be estimated. The incomplete data augmented by some "missing data" are termed the *complete data*, and exist in the complete data space, X, with the probability density $f(x|\Phi)$. There is assumed to be a many-to-one mapping from X into Y such that $g(y|\Phi) = \int_{X(y)} f(x|\Phi)dx$. It was shown in [7] that by iteratively computing the expectation (the E-step) of $Q(\Phi|\hat{\Phi}^{i-1}) = E[\log f(x|\Phi)|y, \hat{\Phi}^{i-1}]$ where i is the iteration number, and then maximizing it (the M-step) such that $\hat{\Phi}^i = \text{argmax}_\Phi Q(\Phi|\hat{\Phi}^{i-1})$, a sequence of estimates is generated such that the corresponding sequence of likelihoods in the incomplete data space is nondecreasing. Under rather easily met conditions [16], this sequence is guaranteed to converge to a maximum. The "missing" data are often chosen to ensure that the maximization in the complete data space can be performed as many one-dimensional maximizations rather than as a single joint maximization over many variables. Lange and Carson [1] made this choice in applying the EM algorithm to the estimation of transmission images in PET. In this section we briefly recapitulate their results in a modified notation. We then expand on them to arrive at the proposed algorithm.

A. The Lange and Carson Algorithm

The scanner geometry is shown schematically in Fig. 1. Assume that the surface of the subject at each projection and at each angle is exposed to a mean number of photons $B_{\theta k}$, where θ is the projection angle and k indexes the transverse position. $B_{\theta k}$ is approximated by the intensity of the accidentals-corrected blank scan at projection angle θ and transverse position k. Assume further that there are enough counts in the blank scan to approximate its intensity by the observed number of counts. For the Lange and Carson model, the observed transmission data are $N_{\theta k}$, and are modeled
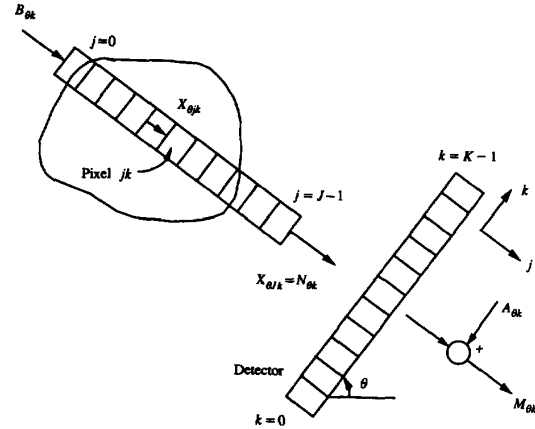

Fig. 1.   Geometry assumed for transmission scans. The observed data would be given by $N_{\theta k}$ if the detectors were ideal, $A_{\theta k}$ are accidental coincidences, and $M_{\theta k}$ are the data that would be observed in the presence of accidental coincidences and finite resolution.

by a Poisson process with intensity $\lambda_{\theta k}$. If one assumes a narrow-beam geometry, this intensity is given by

$$\lambda_{\theta k} = B_{\theta k} e^{-\sum_{j=0}^{J-1} \ell_{\theta jk} \mu_{jk}}. \tag{2}$$

where $\ell_{\theta jk}$ is the length of intersection of pixel jk with the projection line parameterized by θk and μ_{jk}, the pixel value, corresponds to the linear attenuation coefficient at pixel jk. Note that μ_{jk} is assumed to be in a coordinate system rotated by angle θ. This leads to the log-likelihood in the incomplete data space given by

$$\log[g(y|\underline{\mu})] = \sum_\theta \sum_{k=0}^{K-1} \left[-B_{\theta k} e^{-\sum_{j=0}^{J-1} \ell_{\theta jk} \mu_{jk}} \right. \\ \left. -N_{\theta k} \sum_{j=0}^{J-1} \ell_{\theta jk} \mu_{jk} \right] \tag{3}$$

where $N_{\theta k}$ is the number of observed counts at angle θ and transverse position k.

Lange and Carson chose for complete data the observed data augmented by the number of photons incident on each pixel $X_{\theta jk}$, as shown in Fig. 1. This leads to a log-likelihood in the complete data space given by

$$\log f(x|\mu) = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} \sum_\theta \left[(X_{\theta jk} - X_{\theta,j+1,k}) \right. \\ \left. \ln(1 - e^{-\ell_{\theta jk}\mu_{jk}}) - X_{\theta,j+1,k} \ell_{\theta jk} \mu_{jk} \right]. \tag{4}$$

Application of the EM algorithm requires the evaluation of the expression

$$\hat{\underline{\mu}}^i = \text{argmax}_{\underline{\mu}} Q(\underline{\mu}|\hat{\underline{\mu}}^{i-1}) \tag{5}$$
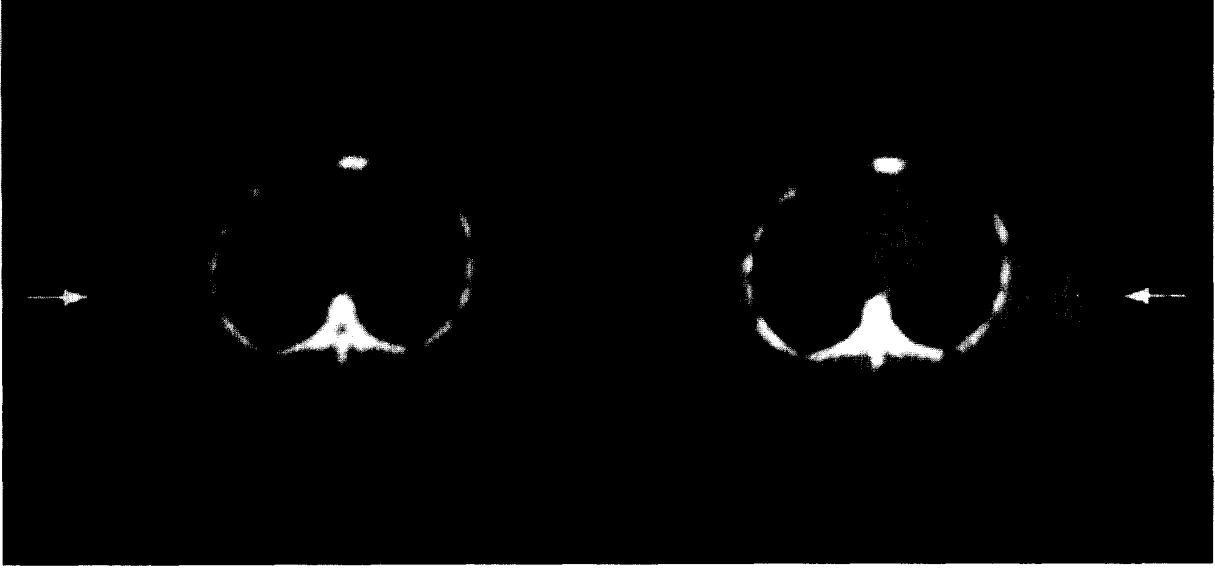
Fig. 2.   Comparison of the true phantom at 8 mm FWHM resolution (left) and the mean of 45 ML reconstructions of the same phantom with two million counts and a 7% accidental coincidence rate (right).

where $\mathcal{Q}(\underline{\mu}|\hat{\underline{\mu}}^{i-1})$ is the conditional mean of the complete data space log-likelihood given by

$$\mathcal{Q}(\underline{\mu}|\hat{\underline{\mu}}^{i-1}) = \sum_{k=0}^{K-1}\sum_{j=0}^{J-1}\sum_{\theta}\left[(\hat{X}_{\theta jk}^{i-1} - \hat{X}_{\theta,j+1,k}^{i-1})\right.$$

$$\left.\ln(1 - e^{-\ell_{\theta jk}\mu_{jk}}) - \hat{X}_{\theta,j+1,k}^{i-1}\ell_{\theta jk}\mu_{jk}\right] \quad (6)$$

$\hat{\underline{\mu}}^{i-1}$ is the estimate of the transmission image at the last iteration, and $\hat{X}_{\theta jk}^{i-1} = E[X_{\theta jk}|N,\hat{\underline{\mu}}^{i-1}]$. Since each term in the inner summation in (6) involves only a single value of μ_{jk}, the Hessian of (6) is a diagonal matrix. This implies that the maximization of $\mathcal{Q}(\underline{\mu}|\hat{\underline{\mu}}^{i-1})$ can be performed as many simple, one-dimensional maximizations rather than as a single, complex one. As shown in [1], evaluating $\hat{X}_{\theta jk}^{i-1}$ leads to the expression

$$\hat{X}_{\theta jk}^{i-1} = \hat{\lambda}_{\theta jk}^{i-1} + N_{\theta k} - \hat{\lambda}_{\theta k}^{i-1} \quad (7)$$

where

$$\hat{\lambda}_{\theta jk}^{i-1} = E[X_{\theta jk}|\hat{\mu}_{jk}^{i-1}] = B_{\theta k}e^{-\sum_{m=0}^{j-1}\ell_{\theta mk}\hat{\mu}_{mk}^{i-1}} \quad (8)$$

and $\hat{\lambda}_{\theta k}^{i-1} = \hat{\lambda}_{\theta Jk}^{i-1}$. The advantage of the EM algorithm over direct maximization of the log-likelihood function is apparent from (6). Maximizing over the log-likelihood in the incomplete space requires the maximization of a transcendental function over n^2 pixels, where n is the dimension of the image. Maximizing the conditional mean of the log-likelihood in the complete data space, on the other hand, consists of n^2 relatively simple maximizations, each over the single variable μ_{jk}.

Equation (6) is maximized in [1] by differentiating it, setting the derivative equal to zero, and solving for $\hat{\mu}^i$. This requires the approximation of a term of the form $1/(e^s - 1)$. Using the first two terms of its Taylor series expansion leads to the

M-step given by

$$\hat{\mu}_{jk}^i = 2\frac{\Delta_{\theta jk}^{i-1}}{\sum_{\theta jk}^{i-1}} \quad (9)$$

where $\Delta_{\theta jk}^{i-1} = \sum_{\theta}\left(\hat{X}_{\theta jk}^{i-1} - \hat{X}_{\theta,j+1,k}^{i-1}\right)$ and $\sum_{\theta jk}^{i-1} = \sum_{\theta}\ell_{\theta jk}$ $\left(\hat{X}_{\theta jk}^{i-1} + \hat{X}_{\theta,j+1,k}^{i-1}\right)$. Using three terms of the Taylor series in the approximation leads to the M-step given by

$$\hat{\mu}_{jk}^i = 2\frac{\Delta_{\theta jk}^{i-1}}{\sum_{\theta jk}^{i-1}}$$

$$+ \frac{2\sum_{\theta}\ell_{\theta jk}^2\left(\hat{X}_{\theta jk}^{i-1}-\hat{X}_{\theta,j+1,k}^{i-1}\right)\left(\sum_{\theta}\left(\hat{X}_{\theta jk}^{i-1}-\hat{X}_{\theta,j+1,k}^{i-1}\right)\right)^2}{3\left(\sum_{\theta}\ell_{\theta jk}\left(\hat{X}_{\theta jk}^{i-1} + \hat{X}_{\theta,j+1,k}^{i-1}\right)\right)^3}.$$

$$(10)$$

B. Modification of the Maximization Step

After implementing the algorithms described by (9), (7), and (10), we found that the images at convergence are smooth versions of the true image, suggesting that the iterates do not converge to the maximum likelihood estimate. We hypothesize that the inexact M-step resulting from the aforementioned approximations no longer increases the likelihood after a certain point. These overly smooth images would result if low-frequency components converge first, as is the case for emission imaging [17].

We approach this problem by using a single step of a stabilized Newton's method to accomplish the M-step. As we shall show, this scheme ensures that the complete data space log-likelihood will increase at every iteration, leading to a generalized EM (GEM) algorithm. Applying a one-dimensional stabilized Newton's method [18] to each term in

(6) leads to an expression of the form

$$\hat{\mu}_{jk}^i = \hat{\mu}_{jk}^{i-1} - a^i \left(\frac{\partial Q(\underline{\mu}|\hat{\mu}_{jk}^{i-1})}{\partial \mu_{jk}} \right) \left(\frac{\partial^2 Q(\underline{\mu}|\hat{\mu}_{jk}^{i-1})}{\partial \mu_{jk}^2} \right)^{-1} \Bigg|_{\mu_{jk}=\hat{\mu}_{jk}^{i-1}}$$

(11)

where $\alpha^i \in [0,1]$ is a search parameter chosen to ensure stability. Note that the derivatives in (11) are scalar functions of μ_{jk}. Taking the first and second derivatives of (6) and substituting into (11) leads to the M-step given by

$$\hat{\mu}_{jk}^i = \hat{\mu}_{jk}^{i-1} + \alpha^i \frac{\sum_\theta \ell_{\theta jk} \left(\hat{\lambda}_{\theta k}^{i-1} - N_{\theta k} \right)}{\sum_\theta \ell_{\theta jk}^2 \hat{\lambda}_{\theta k}^{i-1} \left(e^{\ell_{\theta jk} \hat{\mu}_{jk}^{i-1}} - 1 \right)^{-1}}.$$

(12)

The M-step is stabilized by performing a line-search along the direction specified by the second term in (12) to determine the step size. This is done by choosing α^i from the set $\{1, \frac{1}{2}, \frac{1}{4}, \dots\}$ as follows. For each iteration, we set $\alpha^i = 1$ and compute $\hat{\mu}_{jk}^i$ using (12). We then compute the log-likelihood in the incomplete data space. If it does not increase, we repeat the M-step with α^i set to one-half its previous value. This procedure is repeated until the log-likelihood increases. Since the second derivative in (11) is always negative when evaluated at $\mu_{jk} = \hat{\mu}_{jk}^{i-1}$, the complete data space log-likelihood will be increased for some value of α^i unless the EM algorithm has reached convergence. The statistics used to compute the log-likelihood are the same as those required for the E-step, so if the log-likelihood increases, the next M-step can be performed immediately. Hence, if the line search uses the initial value of α, the M-step proposed here requires only the computation of (12) and a single computation of the log-likelihood.

By Ortega and Rheinboldt [19, Theorem 14.4.3] there is some interval about the maximum, $\hat{\mu}_{jk}^*$, where $\alpha^i = 1$ will result in an increase in log-likelihood of the complete data and the convergence rate will be quadratic, i.e., $|\hat{\mu}_{jk}^* - \hat{\mu}_{jk}^i| \leq \beta |\hat{\mu}_{jk}^* - \hat{\mu}_{jk}^{i-1}|^2$, where β is a constant. By Lange and Carson [1, Lemma 2], the Euclidean distance between successive iterates tends to zero as i tends to ∞, so $\hat{\mu}_{jk}^{i-1}$ will be within this interval for some value of i. Hence, after a sufficient number of iterations, the line-search is unnecessary and the quadratic convergence rate ensure that the result will be near the true maximum. Moreover, the M-step converges to an exact maximizer as the overall algorithm converges.

In our preliminary evaluations presented here and in [5], we used the Lange and Carson M-step for the first twenty iterations to ensure a good starting value for the Newton step. We found that $\alpha^i = 1$ always resulted in an increase of the log-likelihood. Hence, the computationally expensive line-search outlined above is rarely if ever executed in practice, although it is required to ensure convergence.

C. Inclusion of Measurement Error and Accidental Coincidences

Measurement error is incorporated into the algorithm by taking $M_{\theta k}$, as shown in Fig. 1, to be the observed data. Assuming for now the accidental coincidences, $A_{\theta k}$, are zero,

we model $M_{\theta k}$ as a translated Poisson process [20] with intensity $\sum_m p_{k-m} \lambda_{\theta m}$ where p_{k-m} is the probability that a photon incident at detector m is detected at detector k. We now take the incomplete data to be $\{M_{\theta k}\}$. This does not change the complete data space log-likelihood function but does require a modification of the E-step, since we must now compute the conditional means $\hat{X}_{\theta jk} = E[X_{\theta jk}|M, \hat{\underline{\mu}}^{i-1}]$. Using the properties of iterated expectations and (7) yields the expression

$$\hat{X}_{\theta jk}^{i-1} = \hat{\lambda}_{\theta jk}^{i-1} + N_{\theta k}^{i-1} - \hat{\lambda}_{\theta k}^{i-1}$$

where $\hat{N}_{\theta k}^{i-1} = E[N_{\theta k}|M, \hat{\underline{\mu}}^{i-1}]$. This expectation can be evaluated by rewriting $N_{\theta k}$ as the sum of the expected number of counts incident at position θk that were detected at position θm, i.e., as $N_{\theta k} = \sum_m N_{\theta k|m}$, and using the fact that $\Pr[N_{\theta k|m}|M, \hat{\underline{\mu}}^{i-1}]$ is a binomial distribution with $M_{\theta m}$ trials and probability of success $\hat{\lambda}_{\theta k}^{i-1} p_{k-m} / \sum_n \hat{\lambda}_{\theta n}^{i-1} p_{m-n}$. This leads to the result

$$\hat{N}_{\theta k}^{i-1} = \hat{\lambda}_{\theta k}^{i-1} \sum_m \frac{M_{\theta m} p_{k-m}}{\sum_n \hat{\lambda}_{\theta n}^{i-1} p_{m-n}}.$$

(13)

Inclusion of accidental coincidences in the problem adds terms to the log-likelihood, but, as is the case for measurement error, these terms do not depend on the pixel values of the image to be estimated and hence do not affect the M-step. Thus, the effect of accidental coincidences is to require the computation of the conditional mean $E[N_{\theta k}|M, \hat{\underline{\mu}}^{i-1}, \underline{\lambda}^A]$ where $\underline{\lambda}^A$ is the intensity of the accidental coincidence process which we assume to be known. This expectation is evaluated by using the property of iterated expectations to write

$$E[N_{\theta k}|M, \hat{\underline{\mu}}^{i-1}, \underline{\lambda}^A] = E[E[N_{\theta k}|M, \hat{\underline{\mu}}^{i-1}, \underline{\lambda}^A, \underline{A}]|M, \hat{\underline{\mu}}^{i-1}, \underline{\lambda}^A]$$

where $\underline{A}$ is a vector giving the number of actual accidental coincidences at each detector position. Evaluating the inner expectation yields

$$E[N_{\theta k}|M, \hat{\underline{\mu}}^{i-1}, \underline{\lambda}^A]$$
$$= \hat{\lambda}_{\theta k}^{i-1} \sum_m \frac{(M_{\theta m} - E[A_{\theta m}|M, \hat{\underline{\mu}}^{i-1}, \underline{\lambda}^A]) p_{k-m}}{\sum_n \hat{\lambda}_{\theta n}^{i-1} p_{m-n}}.$$

Evaluating the expectation as for (13) leads to $E[A_{\theta m}|M, \hat{\underline{\mu}}^{i-1}, \underline{\lambda}^A] = M_{\theta m} \frac{\lambda_{\theta m}^A}{p_m^* \hat{\lambda}_{\theta m}^{i-1} + \lambda_{\theta m}^A}$, where "$*$" denotes convolution. This yields the expectation step

$$\hat{X}_{\theta jk}^{i-1} = \hat{\lambda}_{\theta jk}^{i-1} - \hat{\lambda}_{\theta k}^{i-1}$$
$$\left(1 - \sum_m \frac{M_{\theta m} p_{k-m}}{\sum_n \hat{\lambda}_{\theta n}^{i-1} p_{m-n}} \cdot \frac{p_m * \hat{\lambda}_{\theta m}^{i-1}}{p_m * \hat{\lambda}_{\theta m}^{i-1} + \lambda_{\theta m}^A} \right)$$

(14)

where $\hat{\lambda}_{\theta k}^{i-1} = \hat{\lambda}_{\theta Jk}^{i-1}$ is the intensity of the $N_{\theta k}$ as given by (2). The M-steps given by (10) and (12) are unchanged. Examination of (14) along with the fact that detector efficiency appears as a common factor in $A_{\theta k}$, $B_{\theta k}$, and $M_{\theta k}$ shows that detector efficiency cancels only if the measurement error is zero, i.e., if p is an impulse.

D. The Narrow-Beam Assumption

Throughout this derivation we have assumed a narrow-beam geometry. This can be justified by noting that a narrow-beam geometry differs from a broad-beam geometry in that no scattered events are detected, i.e., in a narrow-beam geometry every photon that interacts in the subject through Compton scatter is attenuated. This implies that transmission data that either have very low scatter fractions or that have been accurately corrected for scatter are equivalent to data collected in a narrow-beam geometry. Work by Digby et al. [21] shows that the measured linear attenuation coefficient decreases with increasing scatter fraction. Experiments in our laboratory demonstrate that the attenuation correction factor of water as measured with a four hour transmission scan on the Siemens 953B is in error by less than 1%. (This low scatter fraction is due to the use of rod-masking, which only accepts events along lines of response that intersect the known position of the rod source.) In light of [21], we conclude that the fraction of scattered events in the transmission data are quite low, and can safely be approximated as zero for this and similar scanners. Hence we feel that the assumption of a narrow-beam geometry is justified.

In cases where the scatter fraction is significant we propose to treat Compton scatter as an additive Poisson process similar to accidental coincidences, i.e., we propose to correct for a lumped additive noise component consisting of the sum of scattered and accidental coincidences. The intensity of this lumped process would be given by the sum of the intensities of the accidental coincidence and Compton scatter processes.

E. Convergence

Lange and Carson [1] proved convergence for the case where measurement error and accidental coincidences are not present. Their results based on the log-likelihood function in the complete data space apply to the algorithm developed here, since this likelihood is unchanged. The intensity of the measured data is given by

$$\lambda_{\theta k} = \sum_m p_{km} \left(B_{\theta m} e^{-\sum_{j=0}^{J-1} \ell_{\theta j m} \mu_{\theta j m}} \right) + \lambda_{\theta k}^A \qquad (15)$$

which is differentiable in μ. Given these two facts, the convergence proof given in [1] holds for the extensions proposed here.

F. Regularization

Snyder and Miller [14] argue that maximum-likelihood estimates in emission tomography are fundamentally unstable in the sense that the noise in the estimates grows without bound as iterations progress. They have termed this instability the *noise artifact*. Moreover, they argue that the estimation is ill-posed in the sense that finite numerical precision or small errors in the physical model of the scanner can cause large errors in the reconstructed image. They term this second problem the *edge artifact*, since its most obvious manifestation is an overshoot near edges in the reconstructed images. They proposed an elegant solution [14], [15], [20], the *method of*
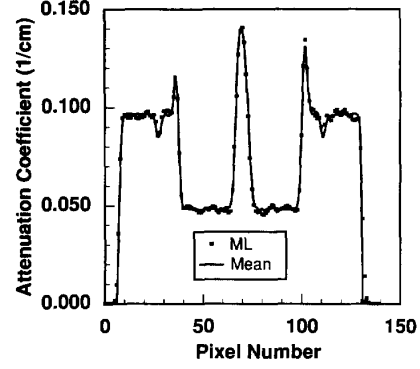

Fig. 3. Profile through the true mean and the sample mean of 45 ML reconstructions of simulated data. The profiles are taken at the level shown by the arrows in Fig. 2.

sieves, that is applicable to the algorithm presented here. This method deals with the noise artifact by constraining the estimate to lie in a suitable subset of the space of nonnegative vectors $S = \{\underline{\mu}: \mu_k = \sum_m \xi_m s_{m|k}\}$ where ξ is an intensity and $\underline{s}$ is the kernel of the sieve. The edge artifact is dealt with by estimating a desired image, $\underline{d}$, that is a low-pass filtered version of the true image, i.e., each pixel is given by $d_k = \sum_m \mu_m r_{m|k}$ where $\underline{r}$ is termed the resolution kernel. This approach can be interpreted as constraining the image to belong to a set of functions for which all frequency components are well-represented by the data.

An appealing feature of the method of sieves is its efficiency: it requires little more computation than the unconstrained algorithm. We implement it by using the algorithm given by (9), (12), and (14) with the following modifications: the measurement error p is modeled by a Gaussian kernel k with a full-width at half-maximum (FWHM) of σ_k; the estimate at each iteration is considered to be the intensity ξ; and the desired image $\underline{d}$ is found by convolving the value of ξ at convergence, ξ^*, with a Gaussian sieve kernel s with FWHM σ_s. The widths of the Gaussian kernels are chosen to satisfy $\sigma_k^2 = \sigma_p^2 + \sigma_s^2 - \sigma_r^2$ [22]. Since we know σ_p^2, the FWHM of the measurement error, only σ_r^2 and σ_s^2 need be chosen. Politte [15] found that choosing $\sigma_r^2 = \sigma_s^2$ yields good results.

III. IMPLEMENTATION

Realization of the algorithm given by (9), (12), and (14) in a manner that is both accurate and efficient requires that choices be made among possible numerical methods. Perhaps the most crucial choice is that of a model for the pixels. Our approach is to model them as point-samples of the underlying function and to approximate the behavior of the function between pixels using bilinear interpolation. The statistics given by (14) are computed over circular regions centered at each pixel. This is accomplished at each angle θ by rotating the estimate of the image at the last iteration, $\hat{\mu}_{jk}^{i-1}$, through θ using bilinear interpolation to a coordinate system in which summing along rows of the image yields a forward-projection at θ. Then the appropriate statistics are computed and rotated through the

Fig. 4.   Sample variance computed from 500 realizations of data for filtered-backprojection (left) and the sample variance computed from 45 realizations for the ML algorithm (right). Each image is scaled to its own maximum. The simulated data contained two million counts and a 7% accidental coincidence rate.

angle $-\theta$. These statistics are accumulated for each projection angle, whereupon the M-step is performed in the coordinate system of the final image. The length of a pixel is assumed to be constant at each angle with a value equal to the pixel size in the final image. Note that the pixels in the rotated coordinate system do not lie at the same positions in the image coordinate system—there is an approximation in the bilinear—interpolation used in the rotation of the statistics into the coordinate system of the final image. With this pixel model, the summation in (2) is equal to a line integral if the underlying image is assumed to vary linearly between pixel centers. This contrasts with the approach of modeling the pixels as square regions of uniform attenuation and calculating the length-of-intersection of the rays and pixels, which would be exact for piecewise-constant images.

The bilinear interpolations have the undesirable effect of smoothing the computed statistics with a triangular filter with a base equal to the width of a pixel. This smoothing can be compensated for by choosing a smaller kernel for the iterations, i.e., make the approximation that $\sigma^2_{k,\mathrm{true}} \approx \sigma^2_k + \sigma^2_\Delta$, and solve for σ^2_k, where $\sigma^2_{k,\mathrm{true}}$ is the effective variance of the kernel used in the iterations, σ^2_k is the FWHM of the kernel $\underline{k}$, and σ^2_Δ is the width of the Gaussian filter that best approximates the triangular smoothing filter of the bilinear interpolation.

The pseudo-code description of the algorithm in the Appendix was implemented in the C language on DECStation 5000 and Sparc 2 workstations. All calculations are done in double precision. The initial algorithm requires approximately 50 sec on a DECStation 5000 to compute one iteration for a scanner having 192 projection angles and 140 transverse samples. No attempt was made to optimize the code for speed. Note that this implementation requires that only one image be exponentiated per iteration.

IV. EVALUATION

The ML algorithm was evaluated in terms of its bias and convergence properties using simulations and phantom studies.

A. Methods

The phantom used in the simulations, shown in Fig. 2, was developed by digitizing tracings of MRI images on a $1,024 \times 1,024$ grid, and then projecting, subsampling by a factor of eight, and smoothing to generate sinograms with an intrinsic resolution of 8 mm FWHM with 4 mm sampling. The arms were generated by adding circles. We assume linear attenuation coefficients of 0. for air, 0.048 cm^{-1} for lung, 0.096 cm^{-1} for tissue, and 0.152 cm^{-1} for bone. Accidental coincidences were simulated as Poisson random variables with constant intensity. The accidentals rate is specified as the ratio of the number of accidental coincidences to the total number of coincidences.

Phantom studies were performed on a Siemens 953B scanner using a Data Spectrum model 2230 elliptical phantom with the Data Spectrum lung insert. The elliptical phantom was filled with water, and the lung insert was filled with shredded Styrofoam and water. The Styrofoam-water mixture had a linear attenuation coefficient of approximately .04 cm^{-1}, with a regional variation depending on how densely the Styrofoam was packed. The cylindrical spine in the phantom had a linear attenuation coefficient of 0.18 cm^{-1}. The phantom (see Fig. 8) was supported in the scanner aperture by a wooden phantom holder constructed from a long piece of lumber with three pieces of plywood (1.59 cm thick) separated by air gaps cradling the phantom.

The Siemens 953B scanner uses three rotating rod sources each of which contained approximately 1.5 mCi of activity. Accidental coincidences were collected in a separate sinogram
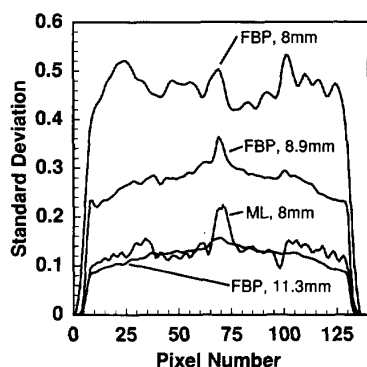
Fig. 5. Profiles through standard deviation images at the level shown in Fig. 4. The standard deviation expressed as a fraction of the linear attenuation coefficient of water is plotted against the pixel position. One set of filtered-backprojection images was reconstructed from unsmoothed data, the others were reconstructed from survival probabilities smoothed with 4 and 8 mm FWHM Gaussian filters. This yielded images with reconstructed resolutions of 8, 8.9, and 11.3 mm, respectively.

and were smoothed with the method proposed by Casey et al. [23] to generate estimates of the accidentals intensity λ_k^A. The rod-masking option, which rejects events that do not lie along lines of response intersecting the rod source, was active. This option reduced the accidental coincidence rate from 7% to 1% and rejected most scattered events. One hour blank scans with accidental coincidences subtracted were used as estimates of the blank intensity. Detector efficiency normalization was not applied. Transmission scans were acquired in the stationary mode, which has a transverse sampling distance of 3.13 mm. We modeled the resolution kernel as a Gaussian with a FWHM of 5.2 mm.

Filtered-backprojection images were computed as follows: the transmission sinogram with randoms subtracted is divided by the scaled blank scan to yield survival probabilities, then a bound is applied to the survival probabilities at projections with zero or negative counts. The bound is computed by thresholding the sinogram to find the maximum path length, assuming that water is present at every pixel along this path, and computing the survival probability for 110% of the estimated path length. The sign-inverted logarithm of the bounded survival probabilities is then reconstructed and scaled.

The ML algorithm was implemented as described in the Appendix. The FWHMs of the sieve and resolution kernels were set to 8 mm for the simulations and to 5.2 mm for the phantom studies.

B. Bias and Variance

An estimate of the bias of the reconstructed transmission images contains three components: bias that is intrinsic to the estimation technique; apparent bias due to approximations made in the implementation; and apparent bias due to the random nature of the sample mean used to compute the bias. Although maximum-likelihood estimators are asymptotically unbiased, no conclusions can be drawn about the bias for finite numbers of counts. Hence the intrinsic bias may be nonzero. The implementation affects the measured bias in two ways: outright errors can lead to erroneous images, and
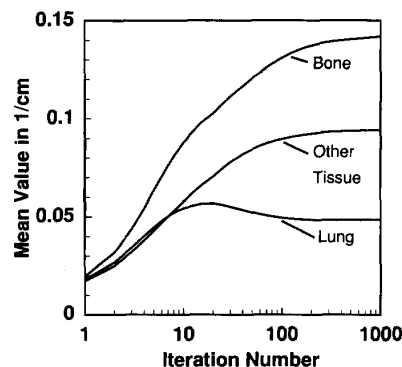

Fig. 6. The sample mean of ML images of the chest phantom averaged over pixels corresponding to bone, lung, and other soft tissue versus iteration. The simulated data contained two million counts and a 7% accidental coincidence rate.
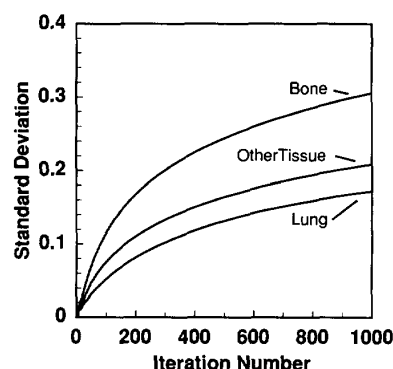

Fig. 7. The sample standard deviation of ML reconstructions of the chest phantom averaged over pixels of bone, lung, and other soft tissue versus iteration. The simulated data contained two million counts and a 7% accidental coincidence rate. The average sample standard deviation is expressed as a fraction of the attenuation coefficient of water.

approximations can cause the estimates to vary from the true maximum-likelihood estimates. Finally, we cannot compute the bias analytically, we must estimate it using a Monte-Carlo simulation, which leads to estimates of the bias that are random.

We have performed a simple evaluation of the bias that consists of computing 45 statistically independent images of the chest phantom with two million counts per data set and a 7% accidentals rate. The algorithm was run for 500 iterations. Pixel-by-pixel estimates of the mean and variance were computed from these images. As shown in Fig. 2, the sample mean of the ML images and the true mean are qualitatively similar. Profiles through the images at the row corresponding to the arrows in Fig. 2 are shown in Fig. 3. The mean of the ML estimates is quite close to the true mean. This suggests that the ML estimates are unbiased, although it is impossible to draw general conclusions from this single case.

The pixel-by-pixel sample variances of ML and FBP reconstructions are shown in Fig. 4. The variance of the filtered-backprojection images is relatively uniform, while the variance of the ML images varies predictably with tissue type: lower variances are associated with pixels with lower means, such as lung; and higher variances are associated with pixels with
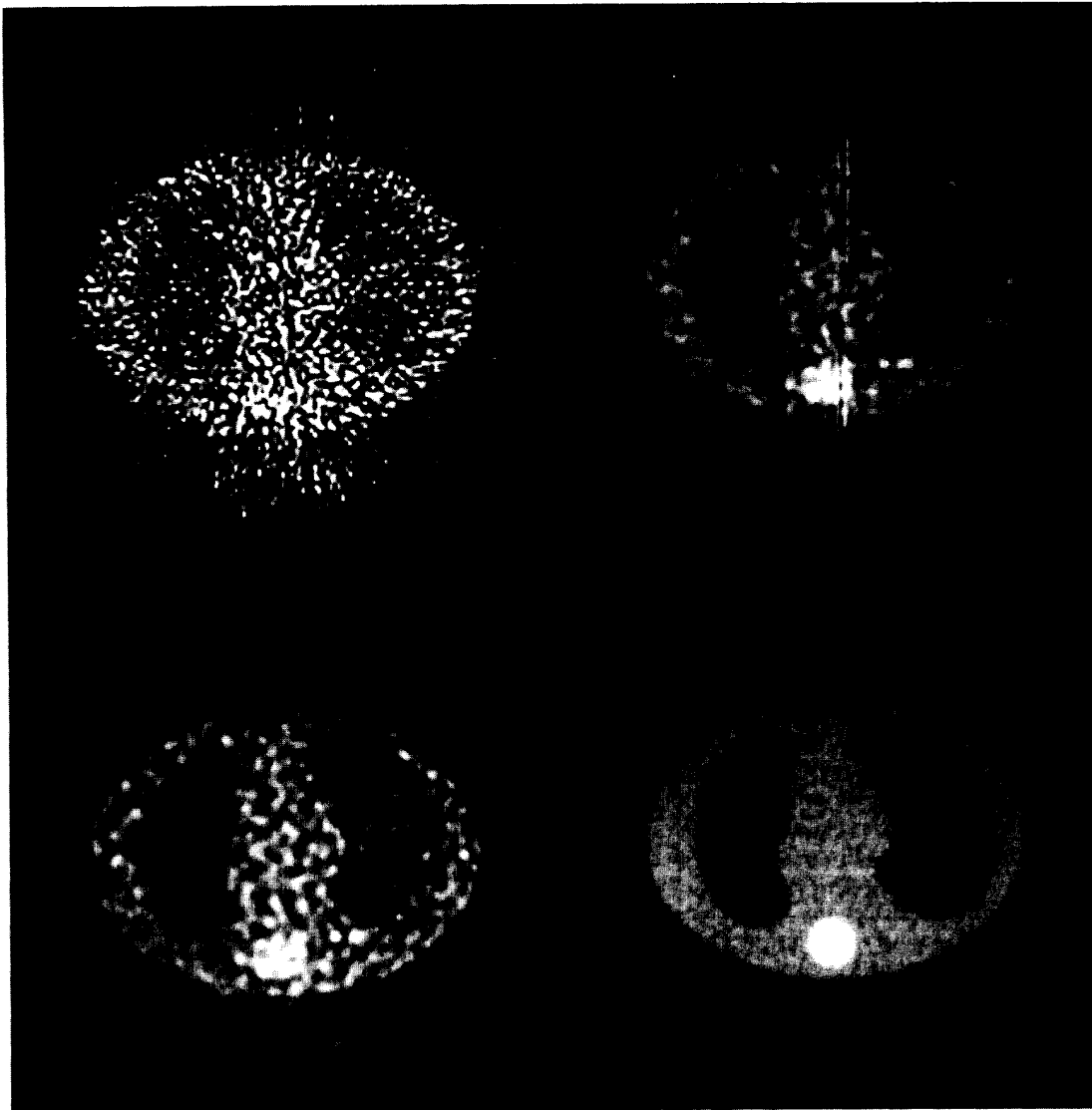
Fig. 8.   Reconstructions of the chest phantom scanned for 2 min on Siemens 953B scanner. Filtered-backprojection with a ramp filter (values less than zero and greater than .2 cm^{-1} were set to 0 and 0.2, respectively) (upper left), filtered-backprojection with a ramp filter and the survival probabilities smoothed with a 5.2 mm FWHM Gaussian filter (upper right), ML reconstruction at the intrinsic resolution (lower left), and a filtered-backprojection reconstruction with a ramp filter of a 240 min scan (lower right).

higher means, such as bone. A quantitative comparison of the variances is given in Fig. 5, where profiles through images of the standard deviation at the row designated by the arrows in Fig. 4 are shown. Three filtered-backprojection variance images were computed: one from unsmoothed data, and two from data smoothed with 4 mm and 8 mm FWHM Gaussian filters. This yielded the variances of images with reconstructed resolutions of 8, 8.9, and 11.3 mm resolutions respectively. In practice, the reconstructed resolution is often 40% greater than the intrinsic resolution, so the variance at 11.3 mm is representative of clinical images. For images reconstructed at

the intrinsic resolution, the ML images have an approximately four-fold improvement in standard deviation. As the reconstructed resolution of filtered-backprojection images becomes larger, i.e., as the images become smoother, the variance of the filtered-backprojection images decreases until, at the smoothest resolution, it is roughly equal to that of the ML images. Hence, we can view the ML algorithm as either producing lower variance images than filtered-backprojection when both are reconstructed to the intrinsic resolution, or as producing an image with better resolution when smoothing filters are selected to deliver the same image variance.

C. Convergence

Simulations were performed to determine the behavior of the mean and variance of image estimates as a function of iteration number. Images were reconstructed from thirty realizations of data, with the sample mean and variance being calculated on a pixel-by-pixel basis for iterates between the first and $1,000^{th}$ iteration. These pixel-by-pixel values were then averaged over tissue types to yield average sample means and variances for bone, lung, and other soft tissue (tissue assigned the same linear attenuation coefficient as water). Pixels on either side of a boundary separating tissue types were excluded from these statistics.

There are two points worth noting about the average sample means plotted in Fig. 6. First, the means converge. Lung and other soft-tissue converged by 250 iterations while bone had not fully converged at 1,000 iterations. Second, pixels corresponding to different tissue types converge at different iterations: lung first, other soft tissue second, and bone last. We conjecture that this is due to the fact that lung is a single, large contiguous region while other soft tissue has small features, and bone consists of many small features. Liow et al. [17] claim that the maximum-likelihood algorithm for emission image reconstruction converges at a rate that varies regionally, with the smallest regions converging last. Our results support this claim.

The increase in image noise as a function of iteration number, i.e., the so-called noise artifact [14], is demonstrated by the plot of the normalized average standard deviation is shown in Fig. 7. The average standard deviation increases monotonically with iteration number regardless of tissue type. We conclude that pixel *values* in individual images do not converge at less than 1,000 iterations regardless of the tissue type. However, the *means* of lung and tissue regions converge with substantially fewer than 1,000 iterations, and bone has nearly converged at 1,000 iterations. This suggests that it is advisable to stop iterating as soon as the means satisfy an appropriate convergence criterion, since this minimizes the image variance.

D. Phantom Studies

The Data Spectrum elliptical phantom with lung inserts was scanned for 2, 10, and 240 min on a Siemens 953B scanner, yielding sinograms with 0.86, 4.3, and 103 million counts. The phantom was supported by a wooden holder. It had a small air bubble at the top. The accidentals fraction was 1%. Filtered-backprojection reconstructions of the 240 min scan yielded estimates of the attenuation coefficient of water of .097 cm^{-1}. This is close enough to the true value to suggest that the scatter fraction was negligible [21].

Images reconstructed from the 2 min scan are shown in Fig. 8. The ML reconstruction was stopped at 250 iterations. Note that the image reconstructed with filtered-backprojection and no smoothing is very noisy and contains streak artifacts along lines of high attenuation. Smoothing the survival probabilities with a Gaussian filter with a FWHM equal to the intrinsic resolution of the scanner (5.2 mm) leads to a much more pleasing image despite prominent streak artifacts, but at a
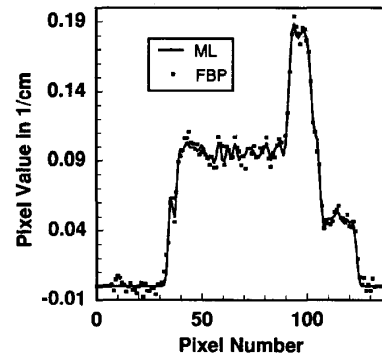

Fig. 9. Vertical profiles through the ML and FBP images in Fig. 11 at the positions indicated by the arrows. The profile was taken vertically such that, going from left to right in the plot, it intersects a small air bubble at the top of the phantom, a large region of water, the simulated spine (which has a linear attenuation coefficient of 0.18 cm^{-1} at 511 keV), and the wooden phantom support.

significant loss of resolution. The ML image reconstructed at the intrinsic resolution is artifact free and qualitatively better than the filtered-backprojection images.

The quantitative accuracy of the algorithm is illustrated by the profiles through ML and filtered backprojection reconstructions of a 240 min scan of the chest phantom shown in Fig. 9. Both images were reconstructed to the intrinsic resolution. The filtered backprojection and ML algorithms track each other accurately, with the filtered-backprojection image being noisier. Assuming that the filtered backprojection algorithm is a standard of accuracy, a plausible assumption given large number of counts in the data, this plot suggests that the ML algorithm yields quantitatively accurate images. We tested the numerical values over a region of water, and found that the average pixel value was 0.0969 cm^{-1}, comparing favorably with the true value of 0.096 cm^{-1}.

A qualitative characterization of the increase of image variance with iteration is given by Fig. 10, which shows several iterates of a single 10 min scan. Although the image variance increases with iteration, the images are not objectionably noisy, even at 2,000 iterations. It appears, therefore, that while the method of sieves does not completely eliminate the noise artifact, it does ameliorate the increase in image variance such that it is acceptable.

Estimates of transmission images reconstructed from a 240 min transmission scan are shown in Fig. 11 at 250 and 2,000 iterations. Note that at 250 iterations the plywood supports are poorly resolved, and that there are small "bump" in the image just above them. Neither of these effects are present in the images at 2,000 iterations. A plot of a single pixel in the leftmost plywood support, indicated by the arrow in Fig. 11, is shown in Fig. 12. The true value of the attenuation coefficient of plywood was found by drawing a large region of interest over the center support in a filtered-backprojection reconstruction. This yielded a value of .049 $+$ $/-$ .001 cm^{-1}. This pixel does not converge until 2,000 iterations, while single pixels selected in the centers of the water, "lung", and "bone" regions had converged by 100 iterations. Since the wood support is the smallest feature in the image, these
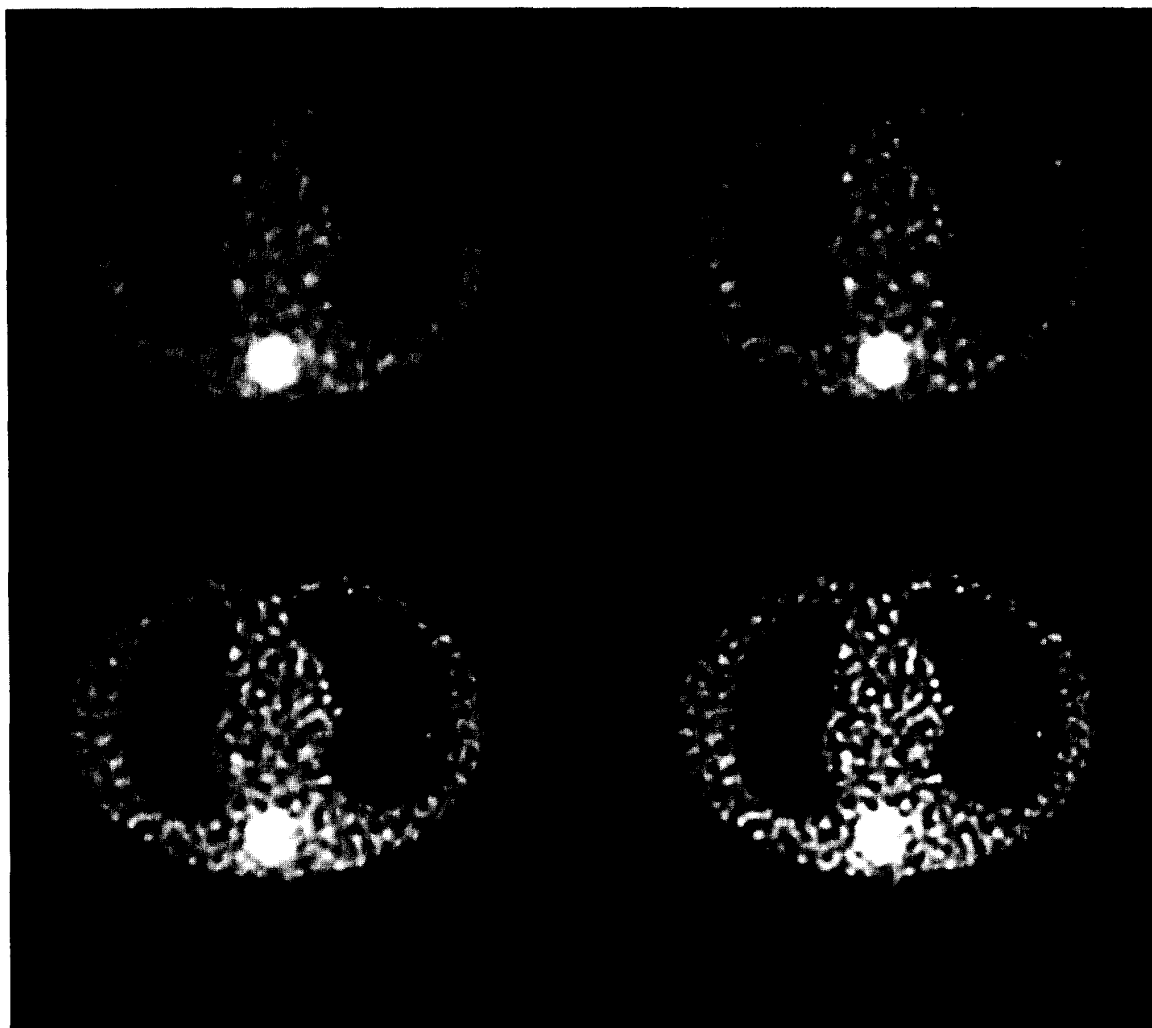
Fig. 10.   ML images reconstructed from a 10 min scan at 250 iterations (upper left), 500 iterations (upper right), 1,000 iterations (lower left), and 2,000 iterations (lower right). The images are scaled to a common maximum.

results suggest that small regions converge more slowly than large regions, confirming the simulations and the result of Liow [17]. This negative conclusion about the ML algorithm is tempered by the fact that an image reconstructed from the same data using filtered-backprojection with a ramp filter overestimated this pixel value by 12.5%, and that prefiltering the data with a 5.2 mm FWHM Gaussian filter resulted in a 5.7% underestimate, the same underestimate as the ML algorithm at the 900th iteration. Hence, the ML algorithm delivers a more accurate estimate at 2,000 iterations and an equally accurate estimate at 900 iterations.

Both the filtered-backprojection and ML reconstructions showed small, barely visible ring artifacts near the "spine" and the air bubble. We were unable to identify any errors in either the filtered-backprojection or ML algorithms, and suspect that the artifacts are due to undersampling of the transmission data. The artifacts are so small that their effect on clinical images would be undetectable.

V. CONCLUSION

We have extended the algorithm proposed by Lange and Carson [1] for computing maximum-likelihood estimates of transmission images in PET to include the effects of measurement error and additive Poisson noise processes such as Compton scatter and accidental coincidences. The algorithm approximates the blank scan intensity with the blank scan, the accidentals intensity by a smoothed, measured accidental coincidence sinogram, and ignores the effects of dead-time and detector efficiency. The M-step is accomplished by using a single iteration of Newton's method at late iterations and the approximation given in [1] at early iterations. Although this M-step is not exact, it does converge to an exact maximizer as the EM algorithm converges. The algorithm is regularized using the method of sieves [14].

The algorithm was implemented and evaluated using both simulation and phantom studies. It requires approximately 50
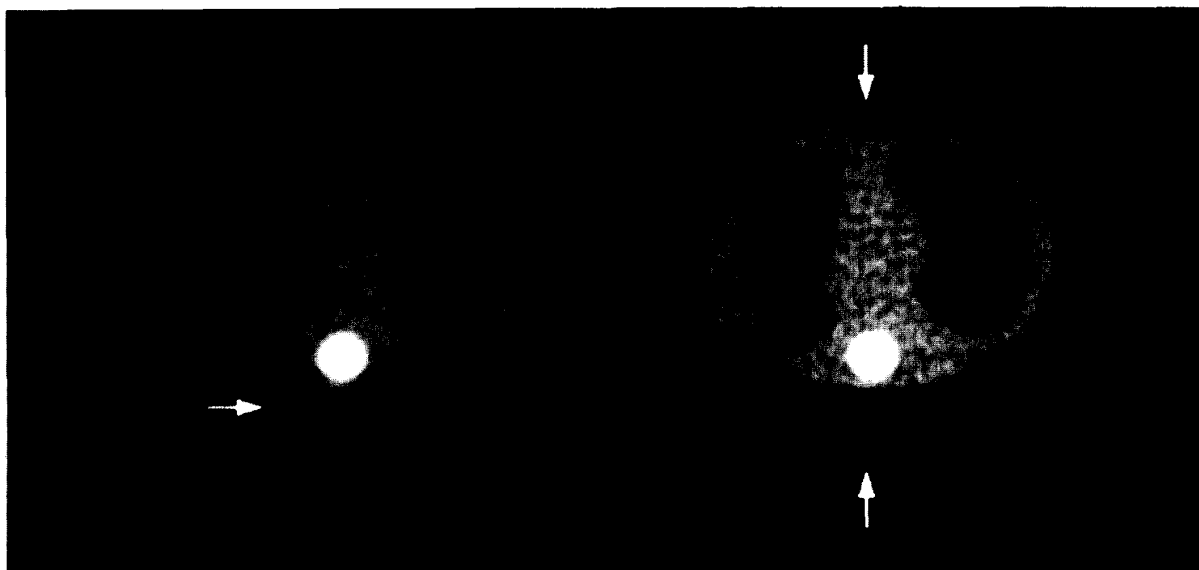
Fig. 11.   ML reconstructions of the chest phantom at 250 iterations (left) and 2,000 iterations (right). The images were reconstructed at the intrinsic resolution.
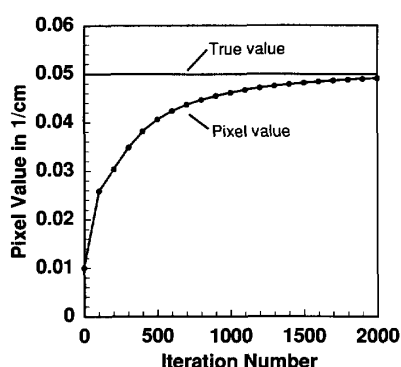

Fig. 12.   A plot of a single pixel (indicated by the arrow) in the smallest feature of the images in Fig. 11 versus iteration number for a 240 min transmission scan.

sec of computation time per iteration on a DECStation 5000. The simulation results suggest that the algorithm computes unbiased estimates, and that their standard deviation is lower than that of filtered-backprojection by approximately a factor of four when images are reconstructed to the intrinsic scanner resolution. Conversely, if smoothing filters are chosen such that the ML and filtered-backprojection algorithms reconstruct images with the same variance, the ML algorithm yields an image with 30% better resolution. Regions of the images converge at different rates, apparently as determined by the sizes of features in these regions. Large features converged in under 200 iterations while small features required in excess of 1,000 iterations. The image variance increases monotonically with iteration, but not to the degree reported for unstabilized algorithms [14].

The simulation results were confirmed with phantom studies. ML reconstructions of a four-hour transmission scan

were accurate and free of large artifacts, although filtered-backprojection and ML images both showed very small, nearly undetectable, ring artifacts near small features in the image. We attribute these small artifacts to undersampling of the data, since they do not appear in reconstructions of simulated data. Short, 2-min scans yield ML images that are obviously superior to the filtered-backprojection images, thereby confirming the simulation results on the relative variance of the images. Pixels in large, homogeneous regions of the four hour scan converged in under 100 iterations while pixels in the smallest feature required approximately 2,000 iterations.

The algorithm's computational complexity poses a major hurdle to its practical use. Several methods have been proposed to accelerate the EM algorithm for emission imaging in PET. Kaufman [24] proposed a line-search method that uses the same search direction as the EM algorithm, but increases the step-size by a scale factor. Although this technique speeds convergence by factors of two to three, it may not be as effective for the transmission algorithm because it requires an evaluation of the log-likelihood function (or equivalently the change in log-likelihood) at each step of the search, and the log-likelihood is more costly to compute for the transmission algorithm than for the emission algorithm.

Another approach explored by Ranganath et al. [25] uses a multigrid approach, i.e., it estimates the image on coarse grids at early iterations, and then refines the grid as the iterations progress. Since the coarse-grid estimates can be computed more quickly, the overall computational cost is decreased. Pan and Yagle [26] showed that in addition to improving the recovery of low-frequency components of the image on coarse grids, multigrid methods also improve the convergence of high-frequency components when the image is locally smooth, i.e., when pixels grouped together in a single grid element are clustered about the same value. However, they also note

that multigrid methods can introduce high-frequency errors in the image, which might actually slow convergence. Our simulation results suggest that the iterations would have to be performed on a fine-grid after the 100th iteration, so between 60% and 95% of the iterations would have to be done on the fine grid if we assume the algorithm requires between 250 and 2,000 iterations for convergence. Hence, an upper bound on the speed improvement would be between a factor of 1.05 and 1.67. Studies by Butler et al. [27] suggest that regularizing with a penalty based on Markov random fields improves the convergence rate of EM algorithms. This might reduce the number of fine-grid iterations enough to make multigrid approaches more effective. We conclude therefore, that a combination of multigrid and line-search methods could improve the convergence rate significantly. Moreover, the time per iteration of 50 sec is based on an inefficient implementation, and could be decreased by a factor of two to four by taking advantage of symmetry conditions and by coding it more efficiently.

APPENDIX

The algorithm as implemented in C is described by the following pseudo-code. All calculations were done in double precision. The pseudo-code uses the following notational conventions: lower case variables refer to quantities in a rotated coordinate system, uppercase variables refer to quantities in the stationary coordinate system, mn indexes images in the stationary coordinate system and ℓk indexes images in a rotated coordinate system. The variable $\chi_{\theta\kappa}$ is equal to the variable $X_{\theta k}$ referred to in the text, except that it is in a rotated coordinate system. The variable ℓ_p is the pixel length and is constant with respect to θ. It corresponds to $\ell_{\theta jk}$ in the text. The variables σ, δ, Σ, and Δ are statistics computed in the expectation step given by (7) and (14).

for($i <$ maximum number of iterations)

$\quad Q_{mn} = e^{-\ell_p \hat{\mu}_{mn}^{i-1}}$ for all pixels mn in the image.

$\quad \Sigma_{mn} = 0$ for all pixels mn.

$\quad \Delta_{mn} = 0$ for all pixels mn.

$\quad \Lambda_{mn} = 0$ for all pixels mn.

$\quad \alpha = 1$.

$\quad$if($0 \le i < 20$) (use Lange and Carson M-step [1], Equations (10) and (14), for first 20 iterations)

$\qquad$for($0 \le \theta < \pi$)

$\qquad\qquad q_{\ell k} = \Re_\theta Q_{mn}$ where $\Re_\theta$ denotes rotation through angle θ using bilinear interpolation for each pixel ℓk in the rotated coordinate system and each pixel mn in the stationary coordinate system.

$\qquad\qquad$for($0 \le k < k_{\max}$) (i.e., each projection k at projection angle θ)

$\qquad\qquad\qquad \lambda_{\ell k}|_{\ell=0} = B_{\theta k}$

$\qquad\qquad\qquad$for($1 \le \ell \le \ell_{\max}$) (i.e., each pixel ℓk along projection k)

$\qquad\qquad\qquad\qquad \lambda_{\ell k} = \lambda_{\ell-1,k} q_{\ell k}$

$\qquad\qquad\qquad\qquad \sigma_{\ell k} = \lambda_{\ell k} + \lambda_{\ell-1,k}$

$\qquad\qquad\qquad\qquad \delta_{\ell k} = \lambda_{\ell k} - \lambda_{\ell-1,k}$

$$\hat{N}_{\theta k} = \lambda_{\ell_{\max} k} \sum_m$$
$$\times \frac{M_{\theta m} p_{k-m}}{\sum_n \lambda_{\ell_{\max} n} p_{m-n}}$$
$$\cdot \frac{p_m^* \lambda_{\ell_{\max} k}}{p_m^* \lambda_{\ell_{\max} k} + \hat{\lambda}_{\theta m}^A}$$

for ($1 \le \ell \le \ell_{\max}$)

$\sigma_{\ell k} = \sigma_{\ell k}$
$\quad + 2(\hat{N}_{\theta k} - \lambda_{\ell_{\max} k})$
$\underline{\Delta} = \underline{\Delta} + \Re_{-\theta} \underline{\delta}$
$\underline{\Sigma} = \underline{\Sigma} + \Re_{-\theta} \underline{\sigma}$
$\hat{\mu}_{mn}^i = \frac{2}{\ell_p} \left(\frac{\Delta_{mn}}{\Sigma_{mn}} + \frac{\Delta_{mn}^3}{3\Sigma_{mn}^3} \right)$

(The M-step given by Equation (10))

else use Newton's method for maximization step. (Equations (14) and (12))

$\quad$for($0 \le \theta < \pi$)

$\qquad q_{\ell k} = \Re_\theta Q_{mn}$

$\qquad$for($0 \le k < k_{\max}$)

$\qquad\qquad \lambda_{\ell k}|_{\ell=0} = B_{\theta k}$

$\qquad\qquad$for ($1 \le \ell \le \ell_{\max}$)

$\qquad\qquad\qquad \lambda_{\ell k} = \lambda_{\ell-1,k} q_{\ell k}$

$$\hat{N}_{\theta k} = \lambda_{\ell_{\max} k} \sum_m$$
$$\times \frac{M_{\theta m} p_{k-m}}{\sum_n \lambda_{\ell_{\max} n} p_{m-n}}$$
$$\cdot \frac{p_m^* \lambda_{\ell_{\max} k}}{p_m^* \lambda_{\ell_{\max} k} + \hat{\lambda}_{\theta m}^A}$$

$\qquad\qquad$for($1 \le \ell \le \ell_{\max}$)

$\qquad\qquad\qquad \delta_{\ell k} = \lambda_{\ell_{\max} k} - \hat{N}_{\theta k}$

$L^{i-1} = L^{i-1} - (p_m^* \hat{\lambda}_{\ell_{\max} k}^{i-1} + \hat{\lambda}^A)$
$\qquad + N_{\theta k} \log\left(p_m^* \hat{\lambda}_{\ell_{\max} k}^{i-1} + \hat{\lambda}^A \right)$
$\underline{\Delta}^{i-1} = \underline{\Delta}^{i-1} + \Re_{-\theta} \delta$
$\underline{\Lambda}^{i-1} = \underline{\Lambda}^{i-1} + \Re_{-\theta} \lambda$

if($L^{i-1} \le L^{i-2}$)

$\quad \underline{\Delta}^{i-1} = \underline{\Delta}^{i-2}$
$\quad \underline{\Lambda}^{i-1} = \underline{\Lambda}^{i-2}$
$\quad \hat{\mu}_{mn}^{i-1} = \hat{\mu}_{mn}^{i-2}$
$\quad \alpha = \frac{\alpha}{2}$

else

$\quad L^{i-2} = L^{i-1}$
$\quad \hat{\mu}_{mn}^i = \hat{\mu}_{mn}^{i-1} - \alpha \frac{(1-p_{mn})}{\ell_p p_{mn}} \frac{\Delta_{mn}^{i-1}}{\Lambda_{mn}^{i-1}}$ (The M-step given by Equation (12).)
$\quad \underline{\Delta}^{i-2} = \underline{\Delta}^{i-1}$
$\quad \underline{\Lambda}^{i-2} = \underline{\Lambda}^{i-1}$
$\quad \hat{\mu}_{mn}^{i-2} = \hat{\mu}_{mn}^{i-1}$

Convolve final estimate with sieve kernel.

REFERENCES

[1] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, no. 2, pp. 306–316, April 1984.
[2] C. A. Pelizzari, C. T. Y. Chen, D. R. Spelbring, R. R. Weichselbaum, and C. T. Chin, "Accurate three-dimensional registration of CT, PET

and/or MR images of the brain," *J. Comput. Assist. Tomogr.*, vol. 13, no. 2 pp. 20–26 1989.

[3] M. R. Palmer, J. G. Rogers, M. Bergstrom., M. P. Beddoes, and B. D. Pate, "Transmission profile filtering for positron emission tomography," *IEEE Trans. Nuclear Sci.,*, vol. 33, pp. 478–481, Feb. 1986.

[4] S. R. Meikle, M. Dahlbom, and S. R. Cherry, "Attenuation correction using count-limited transmission data in positron emission tomography," *J. Nuclear Med.*, vol. 34, no. 1, pp. 143–144, Jan. 1993.

[5] J. M. Ollinger, "The use of maximum *a posteriori* and maximum likelihood transmission images for attenuation correction In PET," *Proc. of the 1992 Nucl. Sci. Symp. Conf. on Medical Imaging*, pp. 1185–1187 (included in Appendix).

[6] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. NS-21 pp. 21–43, June 1974.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statist. Soc.*, B. vol. 39, no. 1, pp. 1–37, 1977.

[8] 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. 6, pp. 106–114, June 1987.

[9] J. A. Browne and T. J. Holmes, "Developments with maximum likelihood x-ray computed tomography," *IEEE Trans. Med. Imaging*, vol. 11, pp. 40–52, Mar. 1992.

[10] K. Sauer and C. Bouman, "Bayesian estimation of transmission tomograms using segmentation based optimization," *IEEE Trans. Nucl. Sci.*, vol. 39, pp. 1144–1152, Aug.1992.

[11] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Sig. Proc.* vol. 41, pp. 534–548, Feb. 1993.

[12] J. A. Fessler, "Segmented attenuation correction for PET using PWLS-ICD," *Conf. Rec. of the 1992 Nucl. Sci. Symp. and Med. Imaging Conf.*, pp. 1182–1184.

[13] E. Mumcuoglu, R. Leahy, and S. Cherry, "A statistical approach to transmission image reconstruction from ring source calibration measurements in PET," *Conf. Rec. of the 1992 Nucl. Sci. Symp. and Med. Imaging Conf.*, pp. 910–913.

[14] D. L. Snyder, M. I. Miller, L. J. Thomas Jr., and D. G. Politte, "Noise and edge artifacts in maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imaging*, vol. 6, pp. 228–238, Sept. 1987.

[15] D. G. Politte and D. L. Snyder, "The use of constraints to eliminate artifacts in maximum likelihood image estimation for emission tomography," *IEEE Trans. Nucl. Sci.*, vol. 35, pp. 608–610, Feb. 1988.

[16] C. F. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statist.*, vol. 11, no. 1, pp. 95–103, 1983.

[17] J. S. Liow and S. C. Strother, "Practical tradeoffs between noise, quantitation, and number of iterations for maximum likelihood-based reconstructions," *IEEE Trans. Med. Imaging*, vol. 10, pp. 563–571, Dec. 1991.

[18] D. G. Luenberger, , *Linear and Nonlinear Programming*. Reading, MA: Addison Wesley, 1984.

[19] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. San Diego, CA: Academic Press, 1970.

[20] D. L. Snyder and M. I. Miller, *Random Point Processes*, Second ed., New York: Springer-Verlag, 1991.

[21] W. M. Digby and E. J. Hoffman, "An investigation of scatter in attenuation correction for PET," *IEEE Trans. Nucl. Sci.*, vol. 36, no. 1, pp. 1038–1042, 1989.

[22] D. L. Snyder and M. I. 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.

[23] M. E. Casey and E. J. Hoffman, "Quantitation in positron emission computed tomography: A technique to reduce noise in accidental coincidence measurements and coincidence efficiency calibration," *J. Comput. Assist. Tomogr.*, vol. 10, no. 5, pp. 845–850, Sept. 1986.

[24] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imaging*, vol. 7, pp. 37–51, March 1987.

[25] M. V. Ranganath, A. P. Dhawan, and N. Mullani, "A multigrid expectation maximization reconstruction algorithm for positron tomography," *IEEE Trans. Med. Imaging*, vol. 7, pp. 273–278, Dec. 1988.

[26] T. S. Pan and A. E. Yagle, "Numerical study of multigrid implementations of some iterative image reconstruction algorithms," *IEEE Trans. Med. Imaging*, vol. 10, pp. 572–588, Dec. 1991.

[27] C. S. Butler and M. I. Miller, "A maximum a posteriori estimation for SPECT using regularization techniques on massively parallel computers," *IEEE Trans. on Med. Imaging*, Mar. 1993.