A Unified Approach to Statistical Tomography Using Coordinate Descent Optimization

Charles A. Bouman, *Member, IEEE,* and Ken Sauer, *Member, IEEE*

***Abstract*—Over the past ten years there has been considerable interest in statistically optimal reconstruction of cross-sectional images from tomographic data. In particular, a variety of such algorithms have been proposed for maximum *a posteriori* (MAP) reconstruction from emission tomographic data. While MAP estimation requires the solution of an optimization problem, most existing reconstruction algorithms take an indirect approach based on the expectation maximization (EM) algorithm. In this paper, we propose a new approach to statistically optimal image reconstruction based on direct optimization of the MAP criterion. The key to this direct optimization approach is greedy pixel-wise computations known as iterative coordinate decent (ICD). We propose a novel method for computing the ICD updates, which we call ICD/Newton–Raphson. We show that ICD/Newton–Raphson requires approximately the same amount of computation per iteration as EM-based approaches, but the new method converges much more rapidly (in our experiments, typically five to ten iterations). Other advantages of the ICD/Newton–Raphson method are that it is easily applied to MAP estimation of transmission tomograms, and typical convex constraints, such as positivity, are easily incorporated.**

I. Introduction

In THE past decade, emission tomography and other photon-limited imaging problems have benefited greatly from the introduction of statistical methods of reconstruction. Unlike the relatively rigid deterministically based methods such as filtered backprojection, statistical methods can be applied without modification to data with missing projections or low signal-to-noise ratios (SNR's). This makes statistical reconstruction methods well suited to emission problems or transmission tomograms of dense materials. The Poisson processes in emission and transmission tomography invite the application of maximum-likelihood (ML) estimation, simply choosing the parameters in the discretized reconstruction that best match the data. However, due to the typical limits in fidelity of data, ML estimates are usually unstable, and have been improved upon by methods such as regularization, maximum *a posteriori* probability (MAP) estimation [1], or the method of sieves [2].

Manuscript received September 2, 1994; revised June 13, 1995. This work was supported by the National Science Foundation under Grant MIP93-00560, by Electricité de France, under Grant P21L03/2K3208/02/EP648, and by an NEC Faculty Fellowship. The associate editor coordinating the review of this paper and approving it for publication was Prof. Andrew E. Yagle.

C. A. Bouman is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: bouman@ecn.purdue.edu).

K. Sauer is with the Laboratory for Image and Signal Analysis, Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA.

Publisher Item Identifier S 1057-7149(96)01800-3.

Both the ML and MAP reconstructions may be formulated as solutions to optimization problems. However, this optimization problem is a formidable numerical task due to both the number of parameters in the estimate (pixels or voxels) and the number of observations (photon-counting measurements). Since the work of Shepp and Vardi [3], the method of choice for finding ML estimates in emission tomography has been the expectation-maximization (EM) algorithm. The EM algorithm is based on the notion of a set of "complete" data which, if available, would make the estimation problem easier. Lange and Carson [4] also formulated the EM method for the transmission problem.

In the Bayesian problem, the EM approach is less simple to apply to emission tomographic reconstruction. This is because the maximization step has no closed form and itself requires the use of an iterative optimization technique. To address this problem, many approaches have been proposed, but all of them approximate the otherwise intractable maximization step. For the case of Gaussian prior densities, Liang and Hart [5], [6], Herman and Odhner [7], and Herman, DePierro, and Gai [8] have modified the EM approach to include Bayesian estimation. A variety of methods have also been proposed for adapting the EM algorithm to more general Markov random field (MRF) priors. These methods include the generalized EM (GEM) algorithm proposed by Hebert and Leahy [9], [10], the one-step-late (OSL) method proposed by Green [11], and a more general form of DePierro's method [12].

The slow convergence of EM is perhaps its greatest disadvantage and is well documented for ML emission tomography [13]. Ollinger used the EM approach of [4] to solve the transmission problem and found that convergence required from 200 to 2000 iterations[14]. For the emission tomography problem, the complete data are usually the number of detections associated with each pixel/detector combination, while the number of photons from each ray entering and exiting each pixel is the corresponding set for transmission. While the use of such large complete data sets appears to simplify the computation of the ML estimate, Fessler and Hero [15] have shown that a large or "informative" complete data space also slows convergence, and have proposed SAGE [16], a collection of methods designed to limit the size of the complete data set with each pixel update in EM reconstruction, with substantially improved convergence.

The similarity of EM iterations to gradient ascent has often been noted [17], [13], [18], and has led to improvements in computational costs and the understanding of convergence properties of the algorithm. Lange and Fessler have recently developed alternative gradient-type optimization approaches with provable convergence for the transmission problem[19]. The preconditioned conjugate gradient approach of Mumcuoğlu *et al.* is also operationally similar to EM, but is designed independently of the EM notion, and is explicitly formulated for the MAP problem in both emission and transmission reconstruction[20].

In this paper, we expand on the work first presented in [21] and take a direct optimization approach to the problem of MAP image reconstruction of emission and transmission tomograms. It is interesting to note that in spite of their relative simplicity, such methods do not seem to have been well investigated for exact statistical image reconstruction with Poisson measurement noise. However, we show that direct optimization is quite tractable for this problem. The result of [15] also suggests that the fastest convergence of an EM-type formulation may be achieved by using the actual observations as the complete data set, which is equivalent to direct optimization of the ML or MAP functionals. Moreover, a direct method without appeal to EM allows one to use a wide range of efficient algorithms to achieve fast convergence. Our approach requires approximately the same amount of *per iteration* computation as traditional EM methods, but has much more rapid convergence, and therefore reduced computation. We also note that some of the methods developed in this paper have recently been studied in [22], and found to compare favorably to alternative approaches.

Our approach to optimization is based on the sequential greedy optimization of pixel (or voxel) values in the reconstruction. This method, which we refer to as iterative coordinate descent (ICD)[1] goes by a variety of other names including iterated conditional modes (ICM) for MAP segmentation [23], and Gauss–Seidel (GS) for partial differential equations [24]. All these algorithms work by iteratively updating individual pixels or coordinates to minimize a cost functional. The ICD method was first applied to least-squares Bayesian tomographic reconstruction by Sauer and Bouman [25]. This work assumed transmission data with a Poisson distribution, but used a Taylor series expansion to derive a quadratic approximation to the exact log likelihood. Fessler has applied this ICD method to a least squares formulation of the emission problem and has shown that underrelaxation methods can speed ICD convergence[26].

The ICD method has a number of important advantages that make it a good choice for direct optimization. First, ICD can be efficiently applied to the log-likelihood expressions resulting from photon-limited imaging systems. In fact, each ICD iteration is similar in nature and computational complexity to an iteration of the EM algorithm; however, we directly attack the posterior probability function rather than the Q function of EM. Second, the ICD algorithm is demonstrated to converge very rapidly (in our experiments, typically five to ten iterations) when initialized with the filtered back projection (FBP) reconstruction. This is not surprising, since the FBP reconstruction is accurate at low spatial frequencies, and in [25] we showed that the ICD method (also known as Gauss–Seidel) has rapid convergence at high spatial frequencies. The third important advantage of the ICD algorithm is that it easily incorporates typical convex constraints and non-Gaussian prior distributions.[2] Positivity is a particularly important convex constraint which can both improve the quality of reconstructions and significantly speed numerical convergence [27]. Non-Gaussian prior distributions are also important since they can substantially reduce noise while preserving edge detail [11], [28], [29], [30].

Our principal result is a new method for computing the ICD updates, which we call ICD/Newton–Raphson. This method is similar to the classical Newton–Raphson root-finding algorithm, but differs from the conventional approach in treating the prior term separately. This is important since the prior term is may be poorly modeled with a quadratic approximation when non-Gaussian image models are used.

Our analysis starts with an approximation of the optimization problem based on a Taylor expansion of the log-likelihood function. This approximation, which was previously developed for the transmission problem [25], is shown to extend equally well to the emission formulation. The Taylor approximation is important for two reasons. It provides a common conceptual framework for both the emission and transmission problems, and it gives insight into the best choice of numerical techniques for the exact optimization problem. We also show that this Taylor series approximation leads to an expression similar to one proposed by Fessler [26] for the modeling of accidental coincidences in emission reconstructions.

While the approximation above may be accurate enough for some high SNR settings, solving the exact ML or MAP problem is our principal concern here. This leads us to the optimization of the exact posterior distribution using the ICD/Newton–Raphson algorithm. The per-iteration computational cost of ICD/Newton–Raphson is found to be similar to that of the EM algorithm, but unlike EM, the ICD/Newton-Raphson algorithm is easily adapted to the Bayesian problem. Finally, we note that a minor modification of the ICD/Newton–Raphson algorithm can be proved to have global convergence [31].

II. Formulation of Statistical Problem

In this section, we will develop the statistical framework for the MAP reconstruction problem for both the emission and transmission case. We will also review the conventional EM approach for later comparison.

For the emission problem, λ is the N dimensional vector of emission rates, Y is the M dimensional vector of Poisson-distributed photon counts, λ_j represents the emission rate from pixel (voxel) j, and P_{ij} is the probability that an emission from pixel j is registered by the ith detector. Thus, according to the standard emission tomographic model, the random vector Y has the distribution

$$\mathcal{P}(Y = y|\lambda) = \prod_{i=1}^{M} \frac{\exp\{-P_{i*}\lambda\}\{P_{i*}\lambda\}^{y_i}}{y_i!} \tag{1}$$

[1] We choose to use the name ICD since it is most descriptive of the algorithmic approach.

[2] We note that convex constraints, such as positivity, may be thought of as a special type of non-Gaussian prior information, and is therefore consistent with the Bayesian problem formulation.

where the matrix P contains the probabilities P_{ij}, and P_{i*} denotes the vector formed by its ith row. This formulation is general enough to include a wide variety of photon-limited imaging problems, and the entries of P may also incorporate the effects of detector response and attenuation. Using (1), the log likelihood may be computed:

$$\text{(emission)}\ \log \mathcal{P}(Y=y|\lambda) = \sum_{i=1}^{M} \left(-P_{i*}\lambda + y_i \log\{P_{i*}\lambda\} - \log(y_i!)\right). \quad (2)$$

The corresponding result for the transmission case is discussed in [25]. In order to emphasize the similarity of the transmission problem, we use the same notation but interpret λ as the attenuation density of a pixel, and y as the observed photon counts. The log likelihood is then given by

$$\text{(transmission)}\quad \log \mathcal{P}(Y=y|\lambda) = \sum_{i=1}^{M} \left(-y_T e^{-P_{i*}\lambda} + y_i(\log y_T - P_{i*}\lambda) - \log(y_i!)\right) \quad (3)$$

where y_T is the photon dosage per ray. Both log-likelihood functions have the form

$$\log \mathcal{P}(Y=y|\lambda) = -\sum_{i=1}^{M} f_i(P_{i*}\lambda) \quad (4)$$

where $f_i(\cdot)$ are strictly convex and differentiable functions. This common form will lead to similar methods of solving these two problems.

For the emission problem, maximum-likelihood (ML) estimation of λ from y yields the optimization problem

$$\hat{\lambda}_{ML} = \arg\min_{\lambda} \sum_{i=1}^{M} (P_{i*}\lambda - y_i \log\{P_{i*}\lambda\}).$$

Probably the most widely applied algorithm for finding $\hat{\lambda}_{ML}$ is expectation-maximization (EM) [32], which was first applied by Shepp and Vardi [3] to the emission tomographic problem. EM solves the ML estimation by hypothesizing the existence of *complete data*, which would allow very simple estimation of λ if available. For the emission tomographic problem, these observations are the number of photons emitted from each discretized cell of the reconstruction region registered at each detector. The iteration resulting from the EM formulation is

$$\lambda_j^{n+1} = \frac{\lambda_j^n}{\sum_{m=1}^{M} P_{mj}} \sum_{i=1}^{M} \frac{y_i P_{ij}}{\sum_{l=1}^{N} P_{il}\lambda_l^n} \quad (5)$$

where n indicates the number of the iteration, updating the entire reconstruction. Because the log likelihood is concave, this approach can be shown to converge to the ML estimate [17].

In Sections III and IV, we will show that the exact ML or MAP reconstruction may also be computed through direct optimization using the ICD algorithm. The ICD algorithm works by sequentially optimizing the log likelihood with respect to each pixel (or voxel) value λ_j. In [25], we introduced

TABLE I
TWO MEASURES OF COMPUTATIONAL COMPLEXITY INDEPENDENT OF HARDWARE/SOFTWARE IMPLEMENTATION. THE FIRST MEASURE IS THE NUMBER OF MULTIPLIES PLUS DIVIDES PER FULL ITERATION OF EACH METHOD. THE SECOND MEASURE IS THE NUMBER TIMES THE P MATRIX IS READ FROM MEMORY. N IS THE NUMBER OF POINTS IN THE IMAGE, M_0 IS THE AVERAGE NUMBER OF NONZERO PROJECTIONS ASSOCIATED WITH EACH IMAGE PIXEL, AND K IS THE AVERAGE NUMBER OF ITERATIONS REQUIRED FOR THE HALF INTERVAL SEARCH

Algorithm	# of multi./div.	# of matrix reads
Filtered Backprojection	M_0N	1
EM (maximum likelihood)	$2M_0N$	2
ICD (Gauss-Seidel) with quadratic approximation	$3M_0N$	2
ICD/Half Interval exact	$3KM_0N$	K+1
ICD/Newton-Raphson exact	$4M_0N$	2

a fast algorithm for implementing ICD in the transmission problem when the log likelihood is approximated by a single quadratic function. This basic ICD algorithm exploits the sparse nature of the projection matrix by maintaining a state vector $\tilde{p}^n = P\lambda^n$ of the projected values.

In this paper, we will investigate three new techniques for applying ICD to emission and transmission MAP reconstruction. The first technique is as in [25], but relies on a new quadratic approximation to the log likelihood derived for the emission problem. The second technique, which we will refer to as ICD/half interval, finds the solution through greedy sequential pixel updates, using a half-interval search to solve the problem exactly at each step. Finally, the method we call ICD/Newton–Raphson (ICD/NR) forms a revised quadratic approximation to the log likelihood at each new pixel update. The ICD/Newton–Raphson method solves the MAP reconstruction problem exactly, with the optimum estimate being its only fixed point. It is this algorithm which we propose as our best solution for the general statistical tomographic reconstruction problem. The details of computation will follow in Sections III and IV.

In order to compare these various algorithms, we will need an objective measure of computational complexity that is independent of the specific hardware or software implementation. For this purpose, we use two figures of merit: the approximate number of multiplies plus divides per full iteration and the number of complete reads of the P matrix. In practice, we have found the number of matrix reads to be a good predictor of algorithm speed since computation is often dominated by memory access time and indexing overhead. Table I lists these two performance measures for the EM algorithm in terms of M_0, the average number of nonzero projection values associated with each pixel.[3] NM_0 is then the number of nonzero entries in the sparse projection matrix P. Notice that one iteration of the EM algorithm requires the computation of two iterations of filtered back projection.

For low signal-to-noise ratio medical imaging problems, the shortcomings of ML estimation are well documented [2], [33], [34]. Therefore, many researchers have resorted to some form of regularized estimation for tomographic inversion. Maximum *a posteriori* probability (MAP) estimation addresses this problem by adding regularization in the form of an *a priori* density for λ. The MAP estimate has been shown to substan-

[3]The entries in Table I assumes that $M_0 >> 1$, the sparse matrix P, is precomputed and stored, and sums independent of the data are precomputed.

tially improve performance in many image reconstruction and estimation problems. In addition, the computation of the MAP estimate is not prohibitively difficult provided that the log of the prior density is a concave function of λ.

A frequent choice for a prior model is the Gaussian MRF, but the quadratic penalty extracted for the Gaussian often causes excessive smoothing of edges. Several prior models have been developed that include desirable edge-preserving properties and maintain concavity in their log prior densities [11], [35], [29,] [30]. Provided we choose one of these models, the MAP problem is also concave, and iterations converge to the global minimum solution. If the log prior is strictly concave, then this minimum is also unique. Any MRF prior model with a convex potential function may be applied to our MAP formulation.

Throughout this paper we will use the generalized Gaussian MRF (GGMRF) model proposed in [29] to illustrate our methods. The GGMRF model has a density function with the form

$$p_\lambda(\lambda) = \frac{1}{z}\exp\left\{-\gamma^q \sum_{\{j,k\}\in C} b_{j-k}|\lambda_j - \lambda_k|^q\right\}$$

where C is the set of all neighboring pixel pairs, b_{j-k} is the coefficient linking pixels j and k, γ is a scale parameter, and $1 \leq q \leq 2$ is a parameter that controls the smoothness of the reconstruction. This model includes a Gaussian MRF for $q = 2$, and an absolute-value potential function with $q = 1$. In general, smaller values of q allow sharper edges to form in reconstructed images.

Prior information may also be available in the form of constraints on the reconstructed solution. We will assume that the set of feasible reconstructions Ω is convex, and in all experiments we will choose Ω to be the set of positive reconstructions. Combining this prior model with the log likelihood expression of (4) yields the expression for the MAP reconstruction:

$$\hat{\lambda}_{MAP} = \arg\min_{\lambda\in\Omega}\left[\sum_{i=1}^{M} f_i(P_{i*}\lambda) + \gamma^q \sum_{\{j,k\}\in C} b_{j-k}|\lambda_j - \lambda_k|^q\right] \tag{6}$$

While the EM algorithm is not difficult to implement or understand in the ML case, it is not directly and simply applicable to MAP estimation when the complete data is taken to be the number of photons associated with each pixel/detector combination. This is because there is no closed-form solution for the maximization step of the iteration. Hebert and Leahy [9], [10] have developed the GEM algorithm to cope with these effects. The GEM algorithm takes the form of coordinate gradient ascent of the MAP EM cost functional with a heuristic step size that can be adjusted to guarantee convergence. DePierro's majorization method for MAP reconstruction is also guaranteed to converge [12]. The OSL method proposed by Green [11] uses an approximate maximization step based on the previous values for neighboring pixels. For all three of these algorithms, computation per iteration is approximately the same as listed for EM in Table I.

III. Computation Of Approximate MAP Estimate

The first step toward efficient direct optimization of (6) will be to develop a quadratic approximation to the log likelihood functions for the emission and transmission problems. This approximation is useful because it will guide the design of efficient optimization techniques for the exact problem, and because it gives important insight into the method and its relationship to existing reconstruction algorithms.

In the Appendix, we compute the first two terms in a Taylor expansion to find a quadratic approximation to the emission log likelihood of (2). In [25], a similar quadratic approximation was derived for the transmission problem. Both approximations have the form

$$\log \mathcal{P}(Y = y|\lambda) \approx -\frac{1}{2}(\hat{p} - P\lambda)^T D(\hat{p} - P\lambda) + c(y) \tag{7}$$

where $\hat{p}$ is a vector of projection measurements, D is a diagonal matrix, and $c(y)$ is some function of the data. For the purposes of MAP estimation $c(y)$ may be ignored since it does not depend on λ. For the emission case, $\hat{p}$, and D are given by[4]

$$\text{(emission)}\hat{p} = y$$
$$D = \text{diag}\{y_i^{-1}\}$$

while for the transmission case they are given by

$$\text{(transmission)}\hat{p}_i = \ln(y_T/y_i) \tag{8}$$
$$D = \text{diag}\{\text{y}_\text{i}\}. \tag{9}$$

The placement and character of the diagonal matrix D gives immediate insight into both problems. For the transmission problem, D more heavily weights those projections that correspond to large photon counts. This is because the large photon counts tend to reduce the variance of the measured transmission projections. However, in the emission problem large photon counts tend to create greater variance. In this case, D is inversely proportional to the measured photon counts.

Fig. 1 compares the exact log-likelihood functions and their quadratic approximations over a range of photon counts. Both plots assume a single projection y_1 and use a 99% confidence interval. In the Appendix, we show that for both transmission and emission cases the approximation error bound is proportional to $\sum_{i=1}^{M} 1/\sqrt{y_i}$ for large y_i. This is consistent with the plots that show better approximations for larger y_i. In practice, the accuracy of this least squares approximation to the log-likelihood function will depend on the dosage (transmission) or emission rates (emission), and the particular application.

Fessler [26] has independently developed a closely related approximation to the log likelihood for reconstruction of PET imagery from data precompensated for accidental coincidences. In this case, the Poisson model is violated since the observed counts of the compensated data can be negative, and

[4]For notational simplicity, we assume that all $y_i > 0$. The Appendix gives a more general quadratic approximation obtained by treating the terms corresponding to $y_i = 0$ separately.

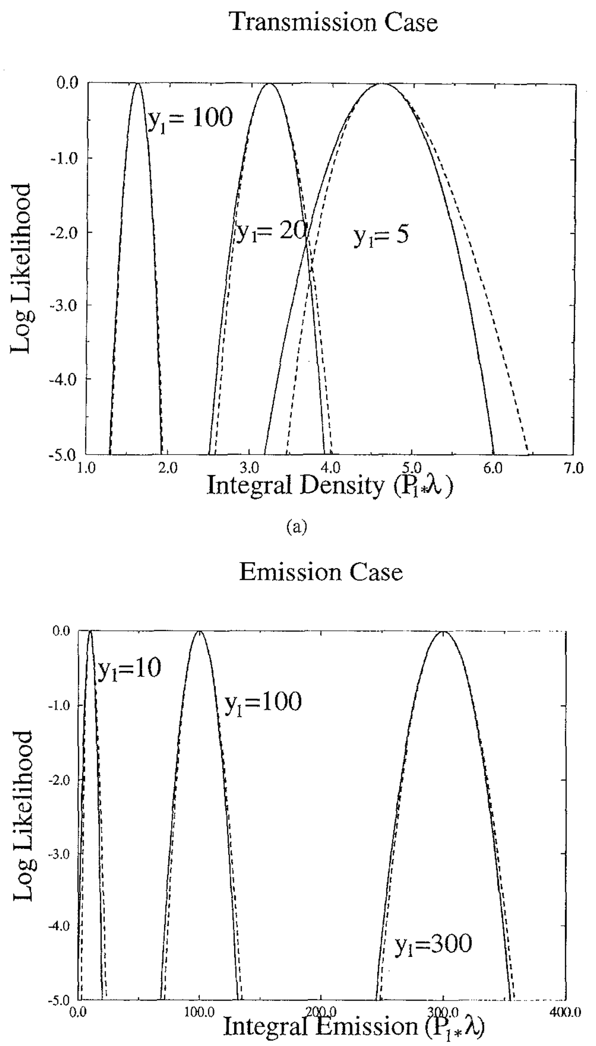


Fig. 1. Plots of the relative log-likelihoods for (a) transmission and (b) emission tomography as a function of a single projection across the reconstructed image. The exact Poisson model (dashed lines) and the quadratic approximation (solid lines) are each plotted for three values of y_1, the observed photon count. In the transmission case, $y_T = 500$. Each of these plots covers a confidence interval of 0.99 for a maximum-likelihood estimate of the projection value.

Fessler employed a Gaussian approximation for the density of corrected counts $\hat{y}$, yielding the objective function

$$\Phi(\lambda) = (\hat{y} - P\lambda)^T \text{diag}\{\sigma_i^{-2}\}(\hat{y} - P\lambda).$$

Here the variance estimates are given by

$$\sigma_i^2 = n_i a_i^2 (a_i^{-1} \tilde{y}_i + 2r_i)$$

where n_i represents detector efficiency, a_i represents attenuation correction factors, $\tilde{y}$ is a smoothed version of $\hat{y}$, constrained to be positive, and r_i are accidental coincidence counts made in an independent measurement. Under this least-squares formulation, Fessler found ICD to converge relatively quickly, improving on the speed of EM iterations. Our approximate formulation of the log likelihood of (9) is similar to Fessler's approximation without accidental coincidences, except for the smoothing of the entries in $\hat{y}$.

IV. Computation of Exact MAP Estimate

To compute the exact MAP reconstruction, we must perform the optimization of (6). Of course, there is a wide variety of techniques from which to choose, but we will use the approximate quadratic structure derived in Section III to guide our selection.

Optimization techniques such as gradient ascent are undesirable because of their slow convergence [25], [36]. Alternatively, conjugate gradient [37] or various preconditioned forms of gradient ascent [38,] [39], [20] techniques may be used since they are known to have rapid convergence for quadratic optimization problems. However, the performance of these techniques is less predictable for the nonquadratic problems resulting from non-Gaussian prior distributions. In fact, in the limit as q approaches 1, the log prior distribution becomes non-differentiable, and gradient based optimization methods can become unreliable [40]. Another significant drawback of conjugate gradient and preconditioning methods is the difficulty of incorporating positivity constraints. The incorporation of positivity constraints for these methods may require the use of "bending" techniques, which are potentially computationally costly [13], [41].

We choose to use the ICD algorithm for a number of reasons. First, it was shown in [25] that the greedy pixel-wise updates of the ICD algorithm produce rapid convergence of the high spatial frequency components in the quadratic problem. Since the FBP can be used as a starting point of the algorithm, the convergence of the low spatial frequencies is less important. Second, the ICD updates work well with non-Gaussian prior models. In fact, the ICM algorithm, which is functionally equivalent to ICD, was developed for the MAP segmentation problem with a discrete prior model [23]. Finally, the ICD algorithm is easily implemented along with convex constraints such as positivity. For this example, each pixel update is simply constrained to be nonnegative.

The ICD algorithm is implemented by sequentially updating each pixel of the image. With each update the current pixel is chosen to minimize the MAP cost function. For emission tomography, the ICD update of the jth pixel is given by

$$\lambda_j^{n+1} = \arg\min_{x \geq 0} \left[\sum_{i=0}^{M} \left[P_{ij}x - y_i \log\left(P_{ij}(x - \lambda_j^n) + P_{i*}\lambda^n\right)\right] + \gamma^q \sum_{k \in \mathcal{N}_j} b_{j-k}|x - \lambda_k^n|^q \right] \tag{10}$$

where $\mathcal{N}_j$ is the set of pixels neighboring j. Notice that in this case λ^n and λ^{n+1} differ at a single pixel, so a full update of the image requires that (10) be applied sequentially at each pixel.

No simple closed-form expression for λ_j^{n+1} results from (10), but there are many optimization techniques that can be employed to find its minimum. We will describe two strategies to the solution of the problem posed by (10). The first strategy is a direct application of half-interval search. However, each iteration of this direct approach is significantly more computationally expensive than an iteration of EM-based methods. The second strategy uses a technique similar to Newton–Raphson search to substantially reduce computation by exploiting the approximately quadratic nature of the log-likelihood function.

Half-interval search may be directly applied to solve (10) by searching for a zero in the derivative of the cost function. The cost function may be analytically differentiated to yield the expression

$$\sum_i P_{ij}\left(1-\frac{y_i}{P_{ij}(x-\lambda_j^n)+P_{i*}\lambda^n}\right) + q\gamma^q \sum_{k\in\mathcal{N}_j} b_{j-k}|x-\lambda_k^n|^{q-1}\mathrm{sign}[x-\lambda_k^n].$$

The disadvantage of this direct approach is that it requires the repeated computation of the derivative. By maintaining a state vector for $P_{i*}\lambda^n$, the derivative may be evaluated with approximately $3M_0$ multiplies and divides. Therefore, the computation required for a complete update is given in Table I as $3KM_0N$ where K is the average number of half-interval iterations required for convergence to the desired precision.

We may reduce computation of ICD updates by exploiting the fact that the log-likelihood function is approximately quadratic. Newton–Raphson optimization works by applying a second-order Taylor series approximation to the function being maximized. For example, if $g(x)$ is the function being maximized and x^n is the current value, then the new value x^{n+1} is given by

$$\begin{aligned} x^{n+1} &= \arg\min_x \left\{ g(x^n)+(x-x^n)g'(x^n)+\frac{1}{2}(x-x^n)^2 g(x^n)\right\} \\ &= x^n - g'(x^n)/g(x^n) \end{aligned}$$

where $g'(x)$ and $g(x)$ are the first and second derivatives of $g(\cdot)$, respectively. Should $g(\cdot)$ be quadratic, we find the exact solution in a single step.

We apply the Newton–Raphson approach to the ICD updates by locally approximating the log-likelihood function as quadratic. However, our method will deviate from conventional Newton-Raphson because we retain the exact expression for the log likelihood of the prior distribution. This is because the prior term is generally not well approximated by a quadratic function. Let θ_1 and θ_2 be the first and second derivatives of the log-likelihood function evaluated for the current pixel value λ_j^n. Using our Newton-Raphson update, the new pixel value is given by

$$\lambda_j^{n+1} = \arg\min_{x\ge 0}\left[\theta_1(x-\lambda_j^n)+\frac{\theta_2}{2}(x-\lambda_j^n)^2 + \gamma^q \sum_{k\in\mathcal{N}_j} b_{j-k}|x-\lambda_k^n|^q\right]. \quad (11)$$

This equation may be solved by analytically calculating the derivative and then numerically computing the derivative's root. The complete set of ICD/Newton–Raphson update equations for emission data is then given by

$$\theta_1 = \sum_{i=1}^{M} P_{ij}\left(1-\frac{y_i}{\tilde{p}_i^n}\right) \quad (12)$$

$$\theta_2 = \sum_{i=1}^{M} y_i\left(\frac{P_{ij}}{\tilde{p}_i^n}\right)^2 \quad (13)$$

$$0 = \theta_1+\theta_2(x-\lambda_j^n)q\gamma^q \cdot \left.\sum_{k\in\mathcal{N}_j} b_{j-k}|x-\lambda_k^n|^{q-1}\mathrm{sign}(x-\lambda_k^n)\right|_{x=\lambda_j^{n+1}} \quad (14)$$

$$\tilde{p}^{n+1} = P_{*j}(\lambda_j^{n+1}-\lambda_j^n)+\tilde{p}^n \quad (15)$$

Equation (14) requires the numerical computation of a root. This can be done using a number of well-known techniques; however, we generally find that a half-interval search works quite well since the function being rooted is monotone decreasing. If the coefficients b_j are nonnegative,[5] then the solution to (14) must satisfy the constraint

$$\min\left\{\lambda_k^n, \lambda_j^n-\frac{\theta_1}{\theta_2} | k\in\mathcal{N}_j\right\} \le \lambda_j^{n+1} \le \max\left\{\lambda_k^n, \lambda_j^n-\frac{\theta_1}{\theta_2} | k\in\mathcal{N}_j\right\}.$$

(The term $\lambda_j^n - \frac{\theta_1}{\theta_2}$ represents the ML value of λ_j under the current state and the local quadratic approximation.) Therefore, these bounds may be used as end points for initiating the half-interval search. Since the half-interval search has guaranteed exponential convergence, one can either choose to terminate after a fixed number of iterations or after a fixed tolerance is reached.

The root-finding operation of (14) is usually computationally inexpensive, since the neighborhood $\mathcal{N}_j$ typically contains only a few pixels.[6] Therefore, the computation is dominated by the $4M_0$ total multiplies and divides required to compute θ_1 and θ_2. A full iteration consists of applying a single Newton–Raphson update to each pixel in λ. This results in $4M_0N$ operations per full image update. In practice, the computation is often dominated by the time required to index through the data. By this measure, the ICD/Newton–Raphson and EM algorithms are computationally equivalent, each requiring two indexings through the projection matrix P.

It should be noted that the distinction between the ICD/Newton–Raphson method and the approximate method of Section III is that for ICD/Newton–Raphson the parameters of the quadratic approximation are recomputed for each new update. This guarantees that the exact MAP reconstruction is the only fixed point of the algorithm.

Recently, we have shown that a small modification in the computation of θ_2 guarantees convergence of the ICD/Newton–Raphson method with any strictly convex prior [31] for both the transmission and emission cases. However, even with the update of (13), we have observed that in all cases

[5] This is required for convexity when $1 < p < 2$.

[6] In some cases, computation time can be reduced by replacing the power function x^{q-1} by a linearly interpolated lookup table.

the convergence is monotone and stable. This is not surprising, since the true log likelihood is very close to quadratic. In fact, ICD/Newton–Raphson method has an intrinsic safety factor since it remains stable in the quadratic case even with overrelaxation by a factor of two [24]. In practice, we have found the ICD/Newton–Raphson method to remain stable even when the over relaxation factor approaches two.

The ICD/Newton–Raphson method may also be applied to the transmission tomography problem. In this case, the parameters θ_1 and θ_2 of (12) and (13) are given by the following equations:

$$\text{(transmission)} \quad \theta_1 = \sum_{i=1}^{M} P_{ij}\left(y_i - y_T e^{-\tilde{p}_i^n}\right)$$
$$\theta_2 = \sum_{i=1}^{M} P_{ij}^2 y_T e^{-\tilde{p}_i^n}.$$

We note that these updates require the evaluation of exponential functions.

V. Experimental Results

Fig. 2 shows the phantoms we used for our emission and transmission tomography experiments together with the FBP reconstructions. The emission phantom and transmission phantoms are of size 64×64 and 128×128, respectively.

The emission phantom in Fig. 2(a) represents higher emission rates with higher image intensity, having zero emission from the background. Rates are scaled to yield a total count of approximately 5×10^4 for the cross-section in Fig. 2(a), with readings taken at 64 equally spaced angles, and 64 perfectly collimated detectors at each angle.

The transmission phantom represents an object of diameter 20 cm and density 0.2 cm^{-1}, with higher density regions of up to 0.48 cm^{-1} added. The dosage per ray (y_T) of only 500 in this experiment results in zero counts at many detectors. Though this is well below typical medical transmission CT imaging rates, low dosage reconstructions are useful in reconstructing approximate attenuation maps for emission imaging [42]. The FBP reconstructions of Fig. 2 show significant noise and streaking artifacts. This is typical of FBP reconstructions, since they do not account for the relative accuracy of projection measurements.

We choose an eight-point neighborhood system for the GGMRF, with normalization of weights $\{b_{j-k}\}$ to a total of 1.0 for each j, and $b_{j-k} = (2\sqrt{2}+4)^{-1}$ for nearest neighbors and $b_{j-k} = (4+4\sqrt{2})^{-1}$ for diagonal neighbors. In order to illustrate the effect of the Bayesian prior, we will compute reconstructions for both $q = 2$ and $q = 1.1$. The first case is equivalent to the common Gaussian prior, and the second does a better job of preserving edges. Since the log prior is strictly concave and differentiable in both cases, convergence of numerical algorithms can be guaranteed. We choose scale parameters yielding the qualitatively best results for comparison of reconstructions under the exact and approximate likelihoods. The parameters of the prior model were $(q = 2.0, \gamma = 1.0)$, $(q = 1.1, \gamma = 3.0)$ for the emission problem, and $(q = 2.0, \gamma = 15.0)$, $(q = 1.1, \gamma = 40.0)$ for the transmission problem. We choose $q = 2.0$ because it results in a Gaussian prior, and we choose $q = 1.1$ because we have found it to give good quality results in a variety of examples. The problem of choosing γ is more complex, since it depends on specific attributes of the reconstructions. Recently, we have developed methods for directly estimating γ [43]–[45]. However, in this example we simply choose values of γ that yield a reasonable visual tradeoff between smoothing and detail preservation.

Figs. 3 and 4 show the results of MAP reconstruction for the emission and transmission problems. In each figure, both the exact MAP reconstructions and the result of the quadratic approximation of the likelihood function in Section III are shown. The exact MAP reconstructions were computed by running the ICD/Newton–Raphson algorithm of Section IV for 150 iterations, at which point changes in both the log *a posteriori* probability and the reconstructed image were negligible. In each case, the cost function being minimized is strictly convex and the algorithm converges to the global minimum. Therefore, the exact reconstruction will be identical to a reconstruction computed using the modified EM algorithm. Since the form of the approximate log likelihood (7) is the same for the emission and transmission problems, the ICD algorithm (called Gauss–Seidel) of [25] was used to calculate both approximate MAP estimates. This algorithm has been shown to converge rapidly and requires approximately $3M_0N$ operations per iterations as listed in Table I.

In both examples of Figs. 3 and 4, there are small but perceptible differences between the exact and approximate reconstructions. Fessler has found that the quadratic approximation can introduce significant bias into transmission reconstructions under low dosages [46]. So exact MAP transmission reconstruction may be of value if precise and absolute density measurements are required.

We concentrate on the exact emission reconstruction problem for comparison of convergence rates to previously proposed algorithms, since it is in this arena that the majority of recent research activity has taken place. (Convergence of ICD in transmission tomography was treated in [25].) Three alternatives to ICD/Newton–Raphson appear in the plots. Green's one-step-late (OSL) algorithm allows EM to be applied to MAP problems by adding a regularizing term to each EM maximization step which is based on the previous iteration's pixel values [11]. The update is made by setting to zero the sum of the gradient of the EM functional and the gradient of the log of the prior density, evaluated at pixel values from the previous iteration. The OSL is very simple to compute, but may fail to converge. The generalized expectation-maximization (GEM) of Hebert and Leahy [9] substitutes an increase in the MAP/EM objective function for the more difficult maximization. GEM features an adjustable step size for the update accompanied by evaluation of the cost functional to guarantee increase in *a posteriori* probability. While the vector $\tilde{p}$ is updated after all pixels have been visited, the pixel values used in evaluating the log of the prior density are updated sequentially. Finally, we include DePierro's method [8], which guarantees convergence through a MAP/EM approach that decouples the computation of pixel

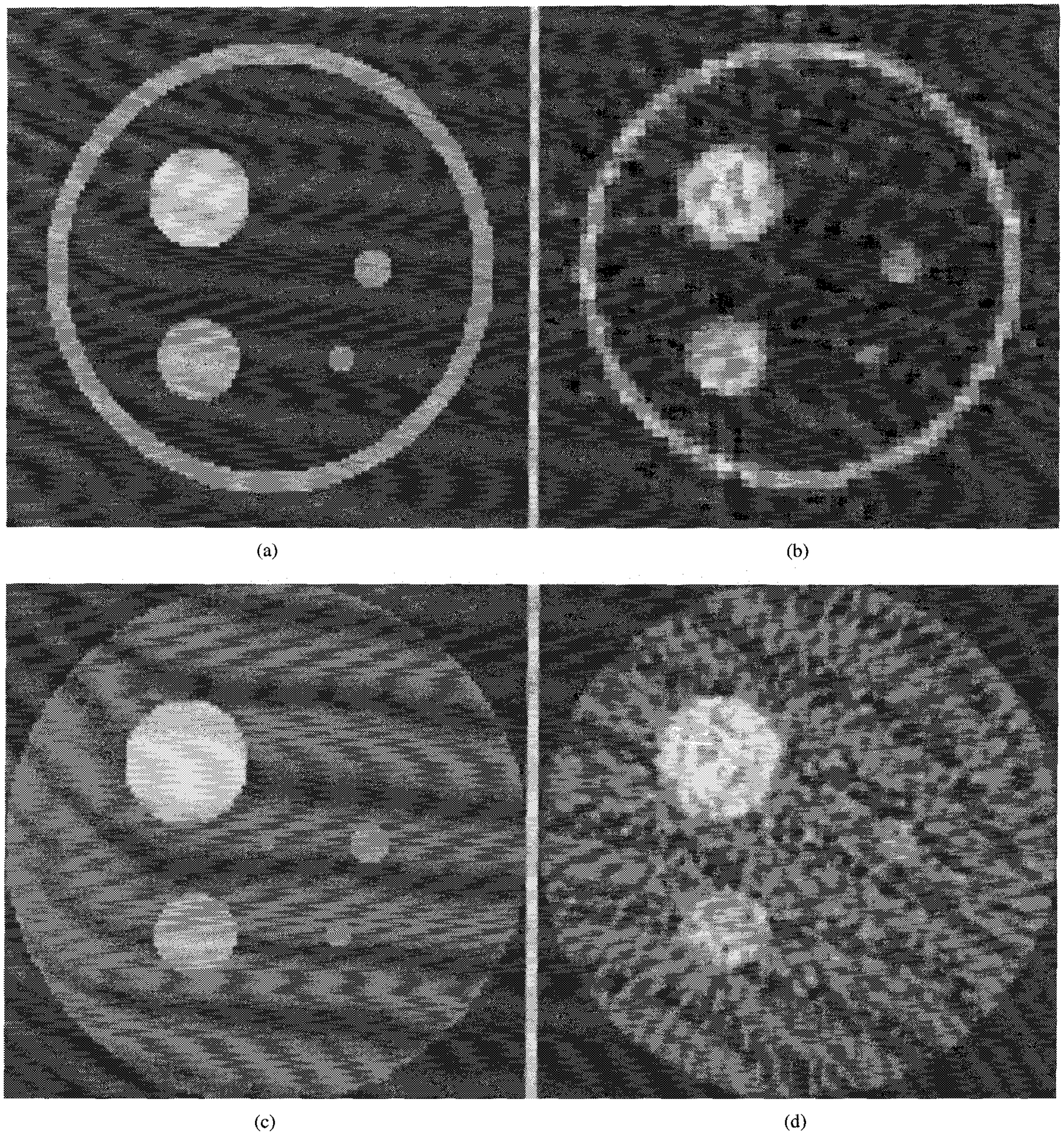


Fig. 2. Original synthetic phantoms and their FBP reconstructions for emission and transmission examples: (a) Emission phantom with higher emission intensities in lighter areas; (b) FBP emission reconstruction; (c) transmission phantom with higher density in lighter areas; (d) FBP transmission reconstruction. FBP reconstructions were computed using a raised cosine rolloff filter, and served as the initial estimate for the iterative statistical methods.

updates in the maximization step. This technique may therefore be applied to complete parallel updates. While DePierro's method was originally designed for the case of a Gaussian prior density, it applies to other convex penalties as well [12], such as the GGMRF with $q = 1.1$.

The three alternative algorithms were implemented without modification to their originally proposed forms. All methods could include a 1-D search for each pixel's update, which is assumed available in both DePierro's method and ICD/NR. Both of these techniques require the minimization of a non-quadratic function at each pixel in the case of a non-Gaussian prior. The adjustable step size of GEM may be replaced by a minimization, and OSL may be augmented to vary the influence of the derivative of the log of the prior at each step to guarantee convergence as well [11].

We will plot convergence performance in terms of complete updates of the image, since like the computational cost measures of Table I, it is independent of implementation. Each of the four methods was initialized in all cases with an FBP reconstruction, which is of negligible cost relative to the ensuing computation. Since the *a posteriori* log likelihood is strictly concave, the solution will not be influenced by this choice of initial condition. Because low-frequency components in the error between the FBP image and the MAP reconstruction will converge most slowly [25], we correct the zero-frequency component of the initial condition with a least-squares estimate

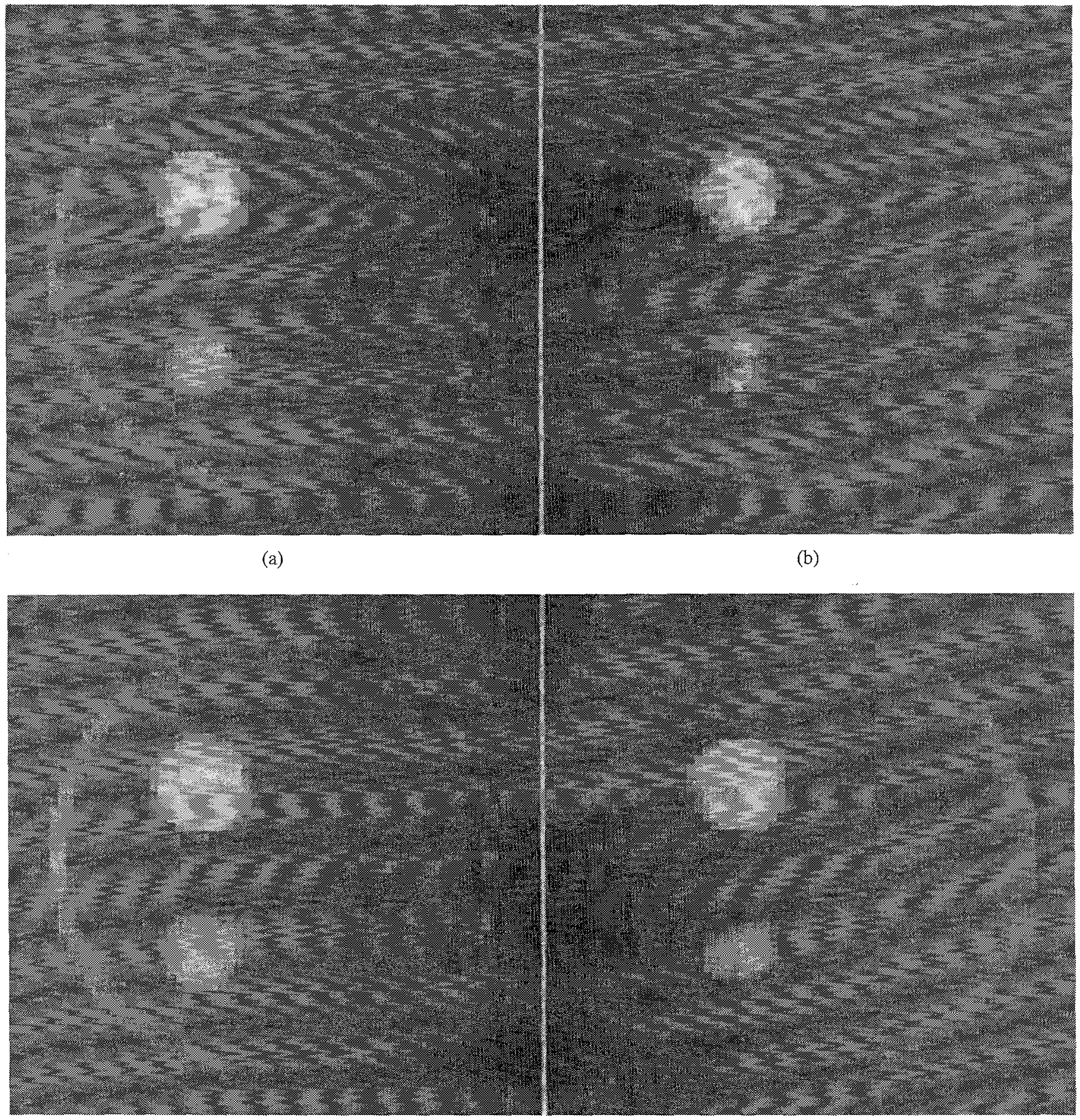


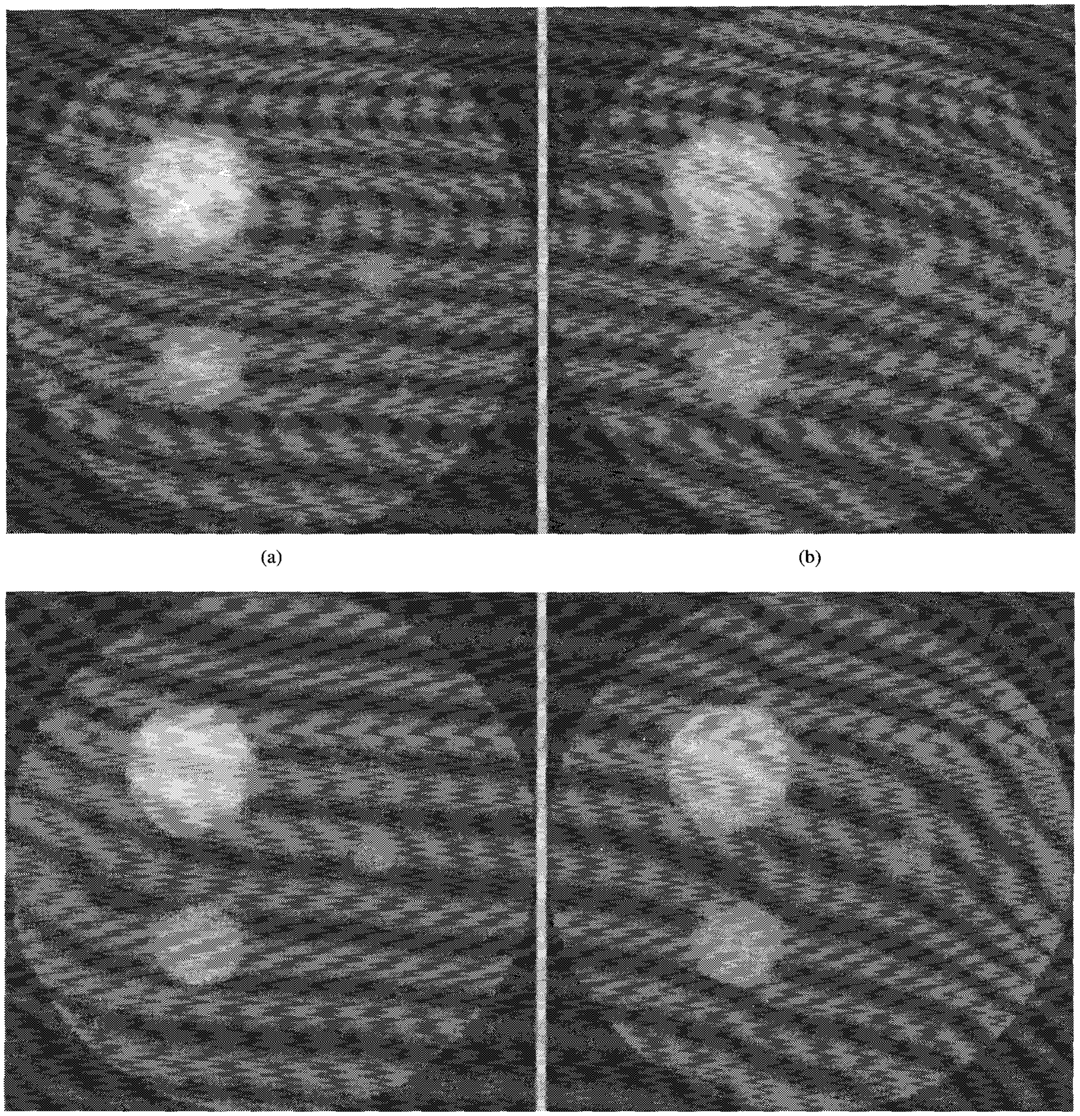
Fig. 3. Emission MAP reconstructions with a Gaussian MRF prior, and (a) exact reconstruction using ICD/Newton–Raphson; (b) quadratic approximation. MAP estimates resulting from GGMRF model with $q = 1.1$; (c) exact reconstruction using ICD/Newton–Raphson; (d) quadratic approximation.

directly from the data. This correction is done by multiplying the FBP image by the appropriate constant.

Fig. 5 shows the convergence rates for the maximum-likelihood problem ($\gamma = 0$). In this case, all three alternative methods reduce to the EM algorithm. The ICD/NR estimate has, for practical purposes, converged after five or six iterations, while EM appears to require over an order of magnitude more. This behavior is at least partly explained by the similarity of EM to gradient ascent, which is particularly slow for this type of problem [25].

Figs. 6 and 7 illustrate similar results for the MAP problem. With the Gaussian prior model, the three EM-based algorithms perform similarly in early iterations, but OSL fails to converge for this case, settling into an oscillation significantly below the maximum *a posteriori* likelihood. For a scale of $\gamma = 2.0$ with the same data, OSL diverged badly from the solution. Both GEM and DePierro's method approach the optimum, but as in Fig. 5, the convergence is much slower than ICD/NR.

Non-Gaussian models allow better preservation of abrupt transitions in MAP estimates. One example that possesses this advantage along with strict convexity of the potential function is the GGMRF with values of q near 1.0. We use $q = 1.1$, which affords good edge preservation with tractable optimization. While this likelihood has first derivatives that are well behaved, the second derivative of the function is unbounded, which may have varying effects on the optimization approaches. Although convergence is similar in this case also, interesting differences exhibit themselves in Fig. 7.



(a) (b)

(c) (d)

Fig. 4. Transmission MAP reconstructions with a Gaussian MRF prior, and (a) exact reconstruction using ICD/Newton–Raphson; (b) quadratic approximation. MAP estimates resulting from GGMRF model with $q = 1.1$; (c) exact reconstruction using ICD/Newton–Raphson; (d) quadratic approximation.

OSL again fails to converge, which is not surprising given the character of the derivative of $|x|^{1.1}$ near the origin. GEM is the fastest of the three EM-type methods in this problem, reaching parity with the ICD/NR solution at about 50 iterations. There is also some potentially interesting asymptotic behavior. After about 60 iterations, when the estimate is undergoing very minor changes, the log likelihood of the GEM estimate slightly exceeds that of ICD/NR. Asymptotic characteristics of ICD/NR in nonlinear problems may require further study and improvement.

VI. Conclusion and Future Directions

We have presented a new method of computing MAP tomographic reconstructions using direct optimization of the log-likelihood function. Each iteration of our proposed ICM/Newton–Raphson algorithm has computation comparable to an iteration of the EM algorithm. However, the new method works well with Bayesian prior distributions and converges much more rapidly than EM. The direct optimization approach also gives a common framework for solving both the emission and transmission tomography problems. The proposed ICM/Newton–Raphson algorithm differs from conventional Newton–Raphson because the prior term is left out of the quadratic approximation. This is important since the prior term may be poorly approximated by a quadratic function.

Experiments indicated that while a fixed quadratic approximation is adequate for some transmission problems, optimization of the exact likelihood appeared to yield improved results in the experiments presented here.

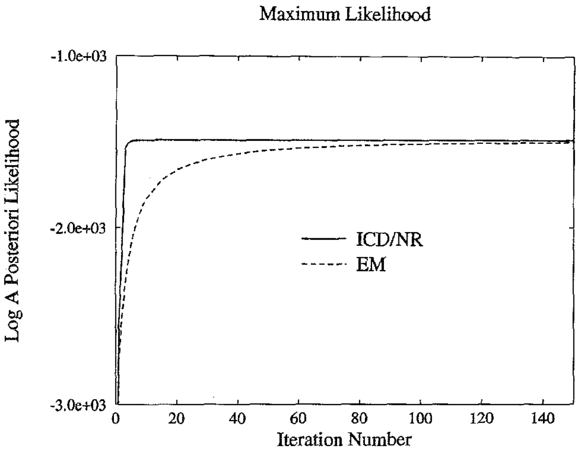


Fig. 5. Convergence of ML estimates using ICD/Newton–Raphson updates and EM. The *a posteriori* likelihood function values are plotted as a function of full iterations.

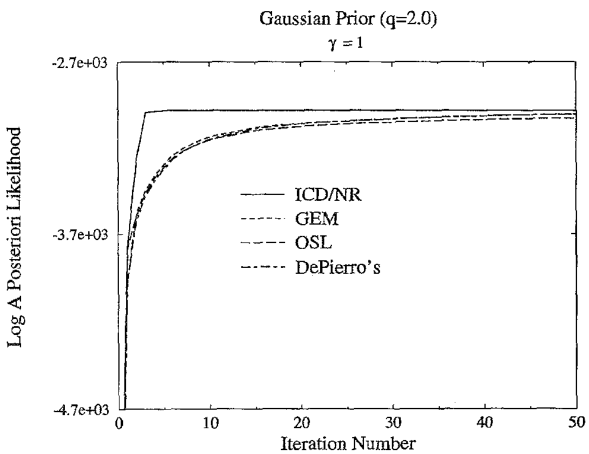


Fig. 6. Convergence of MAP estimates using ICD/Newton–Raphson updates, Green's (OSL), Hebert/Leahy's GEM, and DePierro's method, and a Gaussian prior model with $\gamma = 1.0$

Much work remains to be done to fully understand the use of direct optimization, such as ICD, in tomographic reconstructions. We are currently investigating the effects of over- and underrelaxation to improve convergence speed. Another important modification is the use of simultaneous pixel updates for parallel implementations. Initial investigation indicates that parallel update strategies can be implemented with guaranteed convergence properties [47].

APPENDIX
TAYLOR SERIES APPROXIMATION OF EMISSION LOG LIKELIHOOD

In this Appendix, we derive the Taylor series approximation for the emission log likelihood of (2). We will also determine the convergence behavior of the quadratic approximation for both the transmission and emission case.

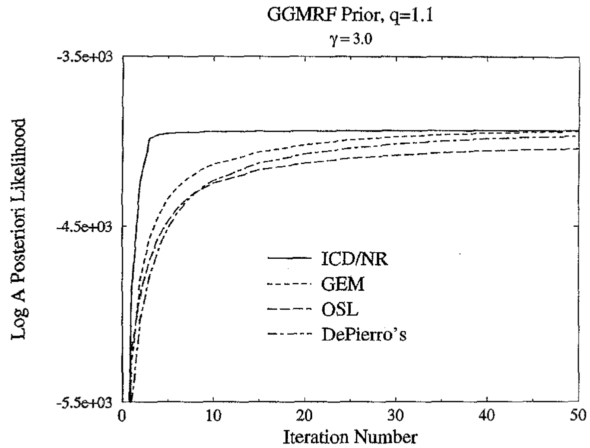


Fig. 7. Convergence of MAP estimates with a generalized Gaussian prior model with $q = 1.1$ and $\gamma = 3.0$.

Using the convention that $\tilde{p}_i = P_{i*}\lambda$, the log likelihood may be expressed as

$$\log \mathcal{P}(Y = y|\lambda) = -\sum_i f_i(\tilde{p}_i)$$

where for the emission case

$$f_i(\tilde{p}_i) = \tilde{p}_i - y_i \log\{\tilde{p}_i\} + \log(y_i!).$$

If we assume that $y_i > 0$, then the gradient of the likelihood function evaluated at $\tilde{p} = y$ has entries

$$\begin{aligned}\left.\frac{\partial \log \mathcal{P}(Y = y|\lambda)}{\partial \tilde{p}_i}\right|_{\tilde{p}=y} &= -1 + \left.\frac{y_i}{\tilde{p}_i}\right|_{\tilde{p}=y} \\ &= 0.\end{aligned}$$

The Hessian is diagonal, with

$$\begin{aligned}\left.\frac{\partial^2 \log \mathcal{P}(Y = y|\lambda)}{\partial \tilde{p}_i^2}\right|_{\tilde{p}=y} &= \left.\frac{-y_i}{\tilde{p}_i^2}\right|_{\tilde{p}=y} \\ &= -\frac{1}{y_i}.\end{aligned}$$

We may account for the terms with $y_i = 0$ by including them separately. Let $S_0 = \{i : y_i = 0\}$ and let $S_1 = S - S_0$. Then the approximation for the log likelihood is given by

$$\begin{aligned}\log \mathcal{P}(Y = y|\lambda) \approx &\sum_{i\in S_1} -\frac{1}{2y_i}(y_i - P_{i*}\lambda)^2 \\ &+ \sum_{i\in S_0} -(y_i - P_{i*}\lambda) + c(y).\end{aligned}$$

If we ignore the terms in S_0, then the log likelihood may be simply written as

$$\begin{aligned}\log \mathcal{P}(Y = y|\lambda) &\approx -\frac{1}{2}(y - P\lambda)^T D(y - P\lambda) + c(y), \\ &\text{with} D = \operatorname{diag}\{y_i^{-1}\}.\end{aligned}$$

We next show that the quadratic approximation for the log-likelihood function converges as $\sum_i 1/\sqrt{y_i} \to \infty$ for both the transmission and emission cases. For the Taylor series approximation of a function $g(x)$, as follows:

$$g(x) = g(a) + g'(a)(x-a) + \frac{g(a)(x-a)^2}{2!} + R_3,$$

we have the Lagrange form of the remainder

$$R_3 = \frac{g^{(3)}(\epsilon)(x-a)^3}{3!} \tag{16}$$

where ϵ is between a and x. We will show in both the transmission and emission tomographic cases that (16) goes to zero on arbitrarily large confidence intervals as the photon counts y_i become large.

In order to cover an arbitrarily large confidence interval for the parameter $\tilde{p}_i$, we will assume that $\tilde{p}_i \in [y_i - k\sqrt{y_i}, y_i + k\sqrt{y_i}]$ where k is any fixed positive integer. For large y_i, the total error is then bounded by

$$\begin{aligned} E &= \sum_i \frac{f_i^{(3)}(\epsilon_i)(\tilde{p}_i - y_i)^3}{3!} \\ &= \sum_i \frac{2y_i}{\epsilon_i^3} \frac{(\tilde{p}_i - y_i)^3}{6} \\ &\leq \sum_i \frac{y_i(k\sqrt{y_i})^3}{3(y_i - k\sqrt{y_i})^3} \\ &\approx \frac{k^3}{3} \sum_i \frac{1}{\sqrt{y_i}}. \end{aligned}$$

Thus the error goes to zero as $\sum_i 1/\sqrt{y_i} \rightarrow 0$.

For the transmission problem

$$f_i(\tilde{p}_i) = y_T e^{-\tilde{p}_i} - y_i(\ln y_T - \tilde{p}_i) + \log y_i!$$

To cover an arbitrary confidence interval, we assume $\tilde{p}_i \in [\hat{p}_i - k/\sqrt{y_i}, \hat{p}_i + k/\sqrt{y_i}]$ where $\hat{p}_i = \log(y_T/y_i)$. This results in the following error bound for large y_i:

$$\begin{aligned} E &= \frac{f_i^{(3)}(\epsilon_i)(\tilde{p}_i - \hat{p}_i)^3}{3!} \\ &= \sum_i y_T e^{-\epsilon_i} \frac{(\tilde{p}_i - \hat{p}_i)^3}{6} \\ &\leq \sum_i \frac{k^3 e^{\frac{k}{\sqrt{y_i}}}}{6\sqrt{y_i}} \\ &\approx \frac{k^3}{6} \sum_i \frac{1}{\sqrt{y_i}}. \end{aligned}$$

Thus, in both cases, the bound on the error magnitude $\rightarrow 0$ as $\sum_i 1/\sqrt{y_i} \rightarrow 0$.

References

[1] S. Geman and D. McClure, "Bayesian image analysis: An application to single photon emission tomography," in *Proc. Statist. Comput. Sect. Amer. Stat. Assoc.*, Washington, DC, 1985, pp. 12–18.

[2] D. Snyder and M. Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," *IEEE Trans. Nucl. Sci.*, vol. NS-32, no. 5, pp. 3864–3871, Oct. 1985.

[3] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-1, no. 2, Oct. 1982.

[4] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomography*, vol. 8, no. 2, pp. 306–316, Apr. 1984.

[5] H. Hart and Z. Liang, "Bayesian image processing in two dimensions," *IEEE Trans. Med. Imag.*, vol. MI-6, no. 3, pp. 201–208, Sept. 1987.

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

[7] G. T. Herman and D. Odhner, "Performance evaluation of an iterative image reconstruction algorithm for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 10, no. 3, Sept. 1991.

[8] G. T. Herman, A. R. DePierro, and N. Gai, "On methods for maximum *a posteriori* image reconstruction with a normal prior," *J. Visual Commun. Image Rep.*, vol. 3, no. 4, pp. 316–324, Dec. 1992.

[9] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 8, no. 2, pp. 194–202, June 1989.

[10] T. Hebert and S. Gopal, "The GEM MAP algorithm with 3-D spect system response," *IEEE Trans. Med. Imag.*, vol. 11, no. 1, pp. 81–90, Mar. 1992.

[11] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–93, Mar. 1990.

[12] A. R. DePierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, no. 1, pp. 132–137, Mar. 1995.

[13] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-6, no. 1, pp. 37–51, Mar. 1987.

[14] J. M. Ollinger, "Maximum-likelihood reconstruction of transmission images in emission computed tomography via the em algorithm," *IEEE Trans. Med. Imag.*, vol. 13, no. 1, pp. 89–101, Mar. 1994.

[15] J. A. Fessler and A. O. Hero, "Complete-data spaces and generalized EM algorithms," in *Proc. ICASSP 1993*, Minneapolis MN, Apr. 27–30, 1993, vol. 5, pp. 1–4.

[16] ——, "Space-alternating generalized EM algorithm for penalized maximum-likelihood reconstruction," Tech. Rep. no. 286, Univ. of Michigan, Feb. 1994.

[17] Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for Poisson emission tomography," *J. Amer. Stat. Assoc.*, vol. 80, no. 389, pp. 8–20, Mar. 1985.

[18] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Trans. Med. Imag.*, vol. MI-6, no. 2, pp. 106–114, June 1987.

[19] K. Lange and J. A. Fessler, "Globally convergent algorithm for maximum *a posteriori* transmission tomography," *IEEE Trans. Image Processing*, vol. 4, no. 10, pp. 1430–1438, Oct. 1995.

[20] E. Ü. Mumcuoğlu, R. Leahy, S. R. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission pet images," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 687–701, Dec. 1994.

[21] C. Bouman and K. Sauer, "Fast numerical methods for emission and transmission tomographic reconstruction," in *Proc. 27th Ann. Conf. Inform. Sci. Syst.*, The Johns Hopkins Univ., Baltimore, MD, Mar. 1993, pp. 611–616.

[22] J. A. Fessler, "Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans," *IEEE Trans. Image Processing*, vol. 4, no. 10, pp. 1439–1450, Oct. 1995.

[23] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, no. 3, pp. 259-302, 1986.

[24] D. M. Young, *Iterative Solution of Large Linear Systems.* New York: Academic, 1971.

[25] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 533–548, Feb. 1993.

[26] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, no. 2, pp. 290–300, June 1994.

[27] P. Oskoui-Fard and H. Stark, "Tomographic image reconstruction using the theory of convex projections," *IEEE Trans. Med. Imag.*, vol. 7, no. 1, pp. 45–58, Mar. 1988.

[28] T. A. Gooley and H. H. Barrett, "Evaluation of statistical methods of image reconstruction through ROC analysis," *IEEE Trans. Med. Imag.*, vol. 11, no. 2, June 1992.

[29] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving map estimation," *IEEE Trans. Image Processing*, vol. 2, no. 3, pp. 296–310, July 1993.

[30] R. Stevenson and E. Delp, "Fitting curves with discontinuities," in *Proc.*

First Int. Workshop Robust Comput. Vision, Seattle, WA, Oct. 1990, pp. 127–136.

[31] Suhail Saquib, Jun Zheng, Charles A. Bouman, and Ken D. Sauer, "Provably convergent coordinate descent in statistical tomographic reconstruction," *IEEE Int. Conf. on Image Processing,* to be presented.

[32] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B,* vol. 39, pp. 1–38, 1977.

[33] E. Veklerov and J. Llacer, "Stopping rule for the mle algorithm based on statistical hypothesis testing," *IEEE Trans. Med. Imag.,* vol. MI-6, pp. 313–319, 1987.

[34] T. Hebert, R. Leahy, and M. Singh, "Fast MLE for SPECT using an intermediate polar representation and a stopping criterion," *IEEE Trans. Nucl. Sci.,* vol. 35, pp. 615–619, Feb. 1988.

[35] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 9, no. 4, pp. 439–446, Dec. 1990.

[36] W. Press, B. Flannery, S. Teukolsky, and Vetterling, *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge, MA: Cambridge Univ. Press, 1988.

[37] F. Beckman, "The solution of linear equations by the conjugate gradient method," in *Mathematical Methods for Digital Computers*, A. Ralston, H. Wilf, and K. Enslein, Eds. New York: Wiley, 1960.

[38] T.-S. Pan, A. E. Yagle, N. H. Clinthorne, and W. L. Rogers, "Acceleration and filtering in the generalized landweber iteration using a variable shaping matrix," *IEEE Trans. Medical Imag.,* vol. 12, no. 2, pp. 278–286, June 1993.

[39] N. H. Clinthorne, T.-S. Pan, P.-C. Chiao, W. L. Rogers, and J. A. Stamos, "Preconditioning methods for improved convergence rates in iterative reconstruction," *IEEE Trans. Med. Imag.,* vol. 12, no. 1, pp. 78–89, Mar. 1993.

[40] K. Sauer and C. Bouman, "Bayesian estimation of transmission tomograms using segmentation based optimization," *IEEE Trans. Nucl. Sci.*, vol. 39, no. 4, pp. 1144–1152, Aug. 1992.

[41] D. M. Goodman, "Appling conjugate gradient to image processing problems," in *Proc. 7th Workshop Multidimensional Signal Processing,* Sept. 1991, Lake Placid, NY, p. 1.1.

[42] R. H. Huesman, S. E. Derenzo, J. L. Cahoon, A. B. Geyer, W. W. Moses, D. C. Uber, T. Vuletich, and T. F. Budinger, "Orbiting transmission source for positron tomography," *IEEE Trans. Nucl. Sci.,* vol. 35, pp. 735–739, Feb. 1988.

[43] C. A. Bouman and K. Sauer, "Maximum likelihood scale estimation for a class of markov random fields," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Adelaide, Australia, Apr. 1994, vol. 5, pp. 537–540.

[44] S. Saquib, C. A. Bouman, and K. Sauer, "Efficient ML estimation of the shape parameter for generalized Gaussians MRF," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Atlanta, GA, May 7–10, 1996, to be published.

[45] S. Saquib, C. A. Bouman, and K. Sauer, "ML parameter estimation for Markov random fields, with applications to Bayesian tomography," *IEEE Trans. Image Processing,* to be published. Also, Tech. Rep. TR-ECE 95-24, Sch. of Elect. and Computer Eng., Purdue Univ., West Lafayette, IN, Oct. 1995.

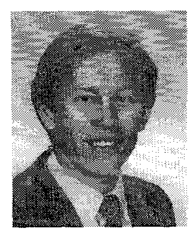
[46] J. A. Fessler, "Sequential iterative algorithms for image reconstruction," presented at the *Midwest Workshop Iterative Image Reconstruction,* St. Louis, MO, June 1994.

[47] K. D. Sauer, S. Borman, and C. A. Bouman, "Parallel computation of sequential pixel updates in statistical tomographic reconstruction," submitted to the *IEEE Int. Conf. Image Processing*, 1995.

Charles A. Bouman (M'89) was born in Philadelphia, PA. He received the B.S.E.E. degree from the University of Pennsylvania, Philadelphia, in 1981 and the M.S. degree in electrical engineering from the University of California at Berkeley, in 1982. In 1987 and 1989, he received the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ.

From 1982 to 1985, he was a staff member in the Analog Device Technology Group at the Massachusetts Institute of Technology, Lincoln Laboratory, Cambridge. In 1989, he joined the faculty of the School of Electrical Engineering at Purdue University, West Lafayette, IN, as an Assistant Professor. His research interests include problems in electronic imaging, image segmentation, analysis, and tomography.

Dr. Bouman is a member of SPIE and IS&T and is currently an associate editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.



Ken Sauer (M'89) was born in Decatur, IN. He received the B.S.E.E. and the M.S.E.E. degrees from Purdue University, West Lafayette, IN, in 1984 and 1985, respectively, and the Ph.D. degree from Princeton University, Princeton, NJ, in 1989 as an AT&T Foundation Fellow.

He is currently an associate professor of electrical engineering at the University of Notre Dame, Notre Dame, IN, and serves as an associate editor for IEEE TRANSACTIONS ON IMAGE PROCESSING. He is involved in research on statistical methods for tomographic image estimation and other nondestructive evaluation problems, and stochastic image modeling.