Newton-style optimization for emission tomographic estimation

Jean-Baptiste Thibault
General Electric Medical Systems
Global Technology Operations, W646
Waukesha, Wisconsin 53188
E-mail: Jean-Baptiste.Thibault@amermsx.med.ge.com

Ken Sauer
University of Notre Dame
Department of Electrical Engineering
Notre Dame, Indiana 46556
E-mail: sauer@nd.edu

Charles Bouman
Purdue University
School of Electrical Engineering
West Lafayette, Indiana 47907-0501
E-mail: bouman@ecn.purdue.edu

Abstract. *Emission computed tomography is widely applied in medical diagnostic imaging, especially to determine physiological function. The available set of measurements is, however, often incomplete and corrupted, and the quality of image reconstruction is enhanced by the computation of a statistically optimal estimate. Most formulations of the estimation problem use the Poisson model to measure fidelity to data. The intuitive appeal and operational simplicity of quadratic approximations to the Poisson log likelihood make them an attractive alternative, but they imply a potential loss of reconstruction quality which has not often been studied. This paper presents quantitative comparisons between the two models and shows that a judiciously chosen quadratic, as part of a short series of Newton-style steps, yields reconstructions nearly indistinguishable from those under the exact Poisson model.* © 2000 SPIE and IS&T. [S1017-9909(00)00502-X]

1 Introduction

Statistical methods of reconstruction are widely applicable in emission tomography and other photon-limited imaging problems. Unlike the relatively rigid, deterministically based methods such as filtered back projection (FBP), they can be used without modification to data with missing projections or low signal-to-noise ratios (SNRs). The Poisson processes in emission and transmission tomography invite the application of maximum-likelihood (ML) estimation. 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, or maximum *a posteriori* probability (MAP) estimation.[1] With the choice of convex potential functions for Markov random field (MRF) style *a priori* image models, both ML and MAP reconstructions may be formulated as large scale convex optimization problems. Many approaches to this challenge have been proposed, among which popular alternatives have been variants of expectation maximization (EM),[2] an approach derived from indirect optimization through the introduction of the notion of an unobservable *complete* data set whose conditional expectations form the algorithmic basis. EM and related methods are called on primarily due to the Poisson likelihood function, while earlier work on least-squares solutions resorted to a variety of more classical numerical methods;[3–6] the motivation for formulating such problems in least squares seems often to have been this availability of simple optimization tools. Linearity of the resulting estimator also opens the problem to more analytical scrutiny. Some previous work[7–9] has considered the differences between the Poisson and least-squares formulations in terms of image qualities in final estimates, but most current work follows the Poisson model.

We study here the visual and quantitative difference between MAP emission tomographic estimates under the Poisson model and under quadratic approximations to the log likelihood. The quadratics are derived from Taylor series expansions of the Poisson log likelihood. As a first goal, we revisit the issue of the viability of a single, fixed approximation based on an expansion before any iterative optimization takes place. Although the diagnostic cost of any loss of quality is not clear from our work, one can

observe that the simplest of quadratics does not always yield a result indistinguishable from the Poisson. We therefore propose a generalization in the form of a short sequence of global quadratic approximations whose minima can be made convergent to the minimum of the Poisson log likelihood. In its limiting case, this technique is a form of Newton iterations in the dimension of the image, and we therefore call the method global Newton (GN). However, as experimental results show, after only a few updates, the approximation yields final image estimates very close to those of the Poisson model. Thus we expect in practice to be able to freeze the approximation after, at most, a handful of updates. Between the refinements of the global quadratic likelihood model, any form of optimization appropriate for the highly coupled equations for the image pixels may be used. This view of, and computational approach to the MAP reconstruction does not necessarily promise great computational savings, since optimization techniques are already in place which converge in few iterations, and which have per-iteration costs similar to those possible within GN. Rather, we hope to show that the likelihood is in practice sufficiently close to good quadratic approximations that choices of numerical methods to solve the Bayesian tomographic inversion problem may be made viewing the optimization as one of solving a large system of linear equations. Should a user have available software packages for quadratic problems, this may allow its application to the emission problem with negligible loss of quality.

While Newton's method would require that the problem formed by each successive approximation be solved exactly before updating the global quadratic, we suggest that a single update of all image pixels suffices to capture most of the per-iteration gain if a relatively fast-converging algorithm is applied. Various enhanced gradient methods, for example, may be well suited to the problem.[10,11] Because it is itself based on *local* quadratic approximations to the Poisson log likelihood and has demonstrated relatively rapid convergence, we apply a form of iterative coordinate descent (ICD) to the pixel optimization phase. Thus we will discuss two cases of quadratic approximation below; the first is global and forms the basis of GN; the second is one dimensional, iteratively solving functions of single pixel values. The latter is used both to compute the exact MAP estimates below and to optimize under the global quadratics of GN.

2 A Global Newton Algorithm for the MAP Reconstruction Problem

2.1 Formulation of the MAP Objective

Statistical image reconstruction requires the evaluation and/or optimization of functionals viewing the probabilistic link between observations and the unknown parameters through the log-likelihood function. For the emission problem, X is the N-dimensional vector of emission rates, Y is the M-dimensional vector of projection data (Poisson distributed photon counts). According to the standard emission tomographic model, the observed photon counts y_i for projection i follow a Poisson distribution with parameter $A_{i*}x$, where A_{i*} is the ith row of the projection matrix A.

Using the convention that $p_i = A_{i*}x$, the log likelihood may be expressed as the sum of strictly convex continuous and differentiable functions

$$\log \mathcal{P}(Y=y|x) = \sum_i \left(y_i \log p_i - p_i - \log(y_i!) \right). \qquad (1)$$

A variety of system nonidealities, such as varying detector sensitivity, scatter under linearized approximations, and attenuation and random coincidences in positron emission tomography (PET) can easily be included in this model with minor modifications. The first three above may typically be incorporated into the transform matrix A, while pre-estimated random coincidence rates are a simple addition to the mean.

Maximum-likelihood (ML) estimation methods are by consensus poor for most emission problems, since high spatial frequency noise tends to dominate ML reconstructions.[12] These high frequencies converge slowly under EM methods of optimization, well after basic image structure is visible. Some ML estimators are therefore ''regularized'' by a uniform initial image estimate and an early termination of the optimization.[13] We formulate the reconstruction instead from the Bayesian point of view, with an explicit *a priori* image model stabilizing the estimator, and optimization methods which are designed to approach the unique maximum of the *a posteriori* probability density as rapidly as possible. Iterations are terminated only when the estimate has stopped evolving more than negligible amounts visually or quantitatively.

We will use throughout this paper the generalized Gaussian MRF (GGMRF)[14] as prior model to illustrate our methods. The GGMRF model has a density function with the form

$$g_x(x) = \frac{1}{z} \exp \left\{ -\frac{1}{q\sigma^q} \sum_{\{j,k\} \in C} b_{j,k} |x_j - x_k|^q \right\},$$

where C is the set of all neighboring pixel pairs, $b_{j,k}$ is the coefficient linking pixels j and k, σ represents the scale of the prior image, and $1 \leq q \leq 2$ is a parameter which 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 Eq. (1) yields the expression of the MAP estimate

$$\hat{x}_{\text{MAP}} = \arg \min_{x \in \Omega} \left[\sum_{i=1}^{M} (p_i - y_i \log p_i) \right.$$

$$\left. + \frac{1}{q\sigma^q} \sum_{\{j,k\} \in C} b_{j-k} |x_j - x_k|^q \right]. \qquad (2)$$

2.2 A Global Newton Method for Poisson Log Likelihood

Direct optimization of Eq. (2) can be simplified somewhat by taking advantage of the approximately quadratic nature of the global log-likelihood function.[15] Dealing directly with the true Poisson during numerical optimization steps is then unnecessary, and efficient optimization can be achieved with a variety of well known methods to compute the descent steps. This is not necessarily motivated by the goal of reducing the number of numerical operations required for the reconstructions; optimization methods exist which converge to the MAP solution rapidly in terms of iteration counts regardless of whether we solve the exact Poisson or a quadratic approximation. These alternative methods can be applied with nearly the same cost to the Poisson likelihood as to the quadratics. But the unboundness of the Poisson log likelihood at the origin, for example, may prove a nuisance, and enforcement of positivity constraints is more easily understood and analyzed under a well-behaved, convex quadratic cost functional. The automatic observation of positivity in pixel values seems to have been an important motivation in the widespread adoption of the EM algorithm for emission tomography. While EM has been accelerated,[16] generalized to the MAP problem,[17] reformulated as a faster sequential algorithm,[18] and speeded (at the cost of reliable convergence) in the ordered subsets EM approach,[19] its derivation and even definition are less well understood by most users than the principles of optimizing a quadratic or, equivalently, solving systems of linear equations. It is our hope that freedom to cogitate on this optimization as a quadratic problem may lead to greater insight into, and greater exploitation of, the unique characteristics of the tomographic inverse problem.

Because edge-preserving *a priori* models may have highly nonquadratic cost functionals, all the work below includes the exact log priors. Most MRF image models add relatively little to the per-iteration computational cost in the tomographic problem, which is dominated by the log-likelihood component. They also serve as stabilizing functionals for the MAP estimator, and will therefore tend to mute, not amplify, the error due to use of the likelihood approximation. In the case of the popular Gaussian MRF, of course, the problem remains entirely quadratic. Though the edge-preserving priors have quantitative advantages, the linearity of the reconstruction as a function of the data in the Gaussian case still may facilitate easier interpretation of errors than the nonlinear case, and we expect that the Gaussian will remain a common choice as prior.

We first compute the leading two terms of a Taylor series expansion of the log-likelihood function in the sinogram domain, where independence of the Poisson photon counts simplifies analysis. Using Eq. (1), the gradient and the diagonal Hessian evaluated at $p = \hat{p}$ have entries

$$\left. \frac{\partial \log \mathcal{P}(Y = y | x)}{\partial p_i} \right|_{p = \hat{p}} = -1 + \frac{y_i}{\hat{p}_i} \tag{3}$$

$$\left. \frac{\partial^2 \log \mathcal{P}(Y = y | x)}{\partial (p_i)^2} \right|_{p = \hat{p}} = -\frac{y_i}{(\hat{p}_i)^2}. \tag{4}$$

Different approaches may be used to take advantage of this approximation. We may evaluate it once before any pixel update at $\hat{p} = y$, using the raw measurement data, and keep this approximation fixed during the convergence process.[15] In this case, an approximation to the log likelihood is

$$\log \mathcal{P}(Y = y | x) \approx - \sum_{i \in S_1} \frac{1}{2 y_i} (y_i - A_{i*} x)^2$$

$$+ \sum_{i \in S_0} A_{i*} x + c(y), \tag{5}$$

where $S_0 = \{i, y_i = 0\}$, $S_1 = S - S_0$. Since this approximation converges to the exact log likelihood as $\Sigma_i y_i^{-1/2}$,[15] we expect the error in the approximation to be far less problematic in high-count data than in low. The FBP image is perhaps the most practical, low-cost starting point for MAP estimation. Therefore one may also profit from an approximation at the point $\hat{p} = A x_{\text{FBP}}$.

A potentially more broadly applicable method involves updating the point of expansion after solving the given quadratic problem, in the manner of Newton's method in N dimensions. Newton's search takes a quadratic approximation to the objective function at each step to compute the next element of the sequential series

$$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1} \cdot \nabla f(x^k). \tag{6}$$

The computationally impractical inversion of the Hessian is often approximated through methods such as preconditioning.[20] Convergence is also locally quadratic but generally not assured for nonconvex objective functions, and improvements on the original idea have been developed to make Newton's method globally robust.[21,22] Applying a similar methodology to the MAP estimation problem in Eq. (2), we may use the previous estimates to globally approximate the log likelihood as quadratic for the next image iteration. The parameters of the Taylor series expansion are then evaluated from the forward projection of the current image estimate. Due to its relation to Newton's method in the dimension of the image, we refer to this algorithm as global Newton (GN). Let $\hat{p}^k$ be the forward projection of the image estimate obtained after optimization on the approximate objective. The expansion point $\hat{p}^k$ is kept fixed for all pixel updates in a stage of GN, then is updated to form the next quadratic likelihood approximation. It is computed from the initial image (typically the FBP reconstruction) for the first iteration. Evaluating Eqs. (3) and (4) at $p = \hat{p}^k$, we obtain a new approximation to Eq. (1)

$$\log \mathcal{P}(Y = y | x) \approx \sum_{i=1}^{M} \left(\frac{y_i}{\hat{p}_i^k} - 1 \right) (y_i - A_{i*} x)$$

$$+ \sum_{i=1}^{M} -\frac{y_i}{2(\hat{p}_i^k)^2} (y_i - A_{i*} x)^2 + c(y). \tag{7}$$

To best exploit the nearly quadratic form of the log likelihood, we seek to update only to a point where the final GN reconstructed image will show no substantial differ-

ence from the exact MAP estimate, terminating the iterations of GN early in the sequence. In Newton's method, the members of this series of quadratic problems each in theory pose a computational load similar to the direct estimator in Eq. (2). Therefore, the complexity of a precise realization of GN appears to be the product of the number of necessary updates of the global quadratic and the cost of each solution. As discussed below, however, a single iteration of an efficient method on a global quadratic appears sufficient to warrant the next step in GN, eliminating this additional cost. Though our implementation does not exactly solve each global quadratic before computing a new expansion, the updates rapidly bring the minimum of the approximate objective close to the exact MAP solution point. The projection estimate $\hat{p}^k$ may then be fixed, and the solution to the final problem form may be refined to essentially complete convergence if desired. The necessary number of updates of $\hat{p}^k$ may be determined by off-line training appropriate to the SNRs typical of the given setting.

2.3 Computation of Estimates via ICD

Under formulations such as Eqs. (2) or (7) large-scale convex optimization must be solved. Such problems are not difficult, but the speed of convergence and simplicity of adaptation to the addition of the regularizing term and positivity constraints in Eq. (2) are important considerations in practice. A variety of techniques may be applied for each GN step, including all classic quadratic (constrained) optimization methods. (As mentioned above, edge-preserving prior models may add a nonquadratic term and may also affect convergence behavior, but typically add minor per-iteration computational cost.) In light of all the above factors, we find types of iterative coordinate descent (ICD) well suited to the problem. 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. The ICD method can be efficiently applied to the log-likelihood expressions resulting from photon-limited imaging systems, is demonstrated to converge very rapidly (in our experiments typically 5–10 iterations) when initialized with the FBP reconstruction, and easily incorporates convex constraints and non-Gaussian prior distributions. While use of approximate second derivative (Hessian) information in this optimization has been introduced in the form of preconditioners for gradient and conjugate gradient algorithms,[10,11,16] ICD's greedy update uses the exact local second derivative directly in one dimension.

Using the exact expression of the emission log likelihood in the MAP estimator, the ICD update of the jth pixel is

$$x_j^{n+1} = \arg\min_{\lambda \geqslant 0} \left\{ \sum_{i=1}^{M} [A_{ij}\lambda - y_i \log(A_{ij}(\lambda - x_j^n) + A_{i*}x^n)] \right.$$

$$\left. + \frac{1}{q\sigma^q} \sum_{k \in \mathcal{N}_j} b_{j,k}|\lambda - x_k^n|^q \right\}, \tag{8}$$

where $\mathcal{N}_j$ is the set of pixels neighboring x_j. In this case x^n and x^{n+1} differ at a single pixel, so a full update of the

image requires that Eq. (8) be applied sequentially at each pixel. Following each pixel update, the forward projections Ax^n are corrected for the difference $x_j^{n+1} - x_j^n$. Since n indexes single pixel updates applied to optimization under a fixed index k for GN approximations, $\{x^k\}$ could be written as a subsequence of $\{x^n\}$.

Rather than solving the exact equation for each pixel resulting from Eq. (2), the ICD/Newton–Raphson (ICD/NR) algorithm[15] exploits the approximately quadratic nature of the log likelihood to reduce computation time. It uses a technique similar to the Newton–Raphson search by *locally* applying a second order Taylor series expansion of the log likelihood as a function of the single pixel value. We retain, however, the exact expression for the prior distribution, because the prior term is often not well approximated by a quadratic function. Should this function be quadratic, we find the exact solution in a single step. We emphasize that this *local* quadratic approximation for each pixel update is a separate process from the *global* approximation to the log likelihood of the GN approach. Let θ_1 and θ_2 be the first and second derivatives of the log likelihood evaluated for the current pixel value x_j^n. Using the Newton–Raphson-type update, the new pixel value is

$$x_j^{n+1} = \arg\min_{\lambda \geqslant 0} \left\{ \theta_1(\lambda - x_j^n) + \frac{\theta_2}{2}(\lambda - x_j^n)^2 \right.$$

$$\left. + \frac{1}{q\sigma^q} \sum_{k \in \mathcal{N}_j} b_{j-k}|\lambda - x_k^n|^q \right\}. \tag{9}$$

This equation may be solved by analytically calculating the derivative and then numerically computing the derivative's root. With p_i^n as the current forward projections, the parameters for the update equations for emission data are then

$$\theta_1 = \sum_{i=1}^{M} A_{ij}\left(1 - \frac{y_i}{p_i^n}\right), \tag{10}$$

$$\theta_2 = \sum_{i=1}^{M} y_i \left(\frac{A_{ij}}{p_i^n}\right)^2. \tag{11}$$

We simply choose a half-interval search to solve Eq. (9) since the function being rooted is monotone decreasing. A full iteration consists of applying a single Newton–Raphson update to each pixel in x. We have observed that in all cases, the convergence of ICD/NR is stable. In fact, we have shown that a small modification in the computation of θ_2 guarantees the global convergence of the method with any strictly convex prior.[23]

Using an approximation such as GN, with Eq. (7) in place of the exact Poisson log likelihood in Eq. (2), the computation of the update of the jth pixel is realized as in Eq. (9) with new expressions for θ_1 and θ_2

$$\theta_1 = \sum_{i=1}^{M} A_{ij}\left(1 + \frac{y_i}{\hat{p}_i^k}\left(\frac{p_i^n}{\hat{p}_i^k} - 2\right)\right), \tag{12}$$

$$\theta_2 = \sum_{i=1}^{M} y_i \left(\frac{A_{ij}}{\hat{p}_i^k} \right)^2. \tag{13}$$

We label this form of coordinate descent ICD/GN. Updating this quadratic approximation at each pixel (replacing $\hat{p}_i^k$ with p_i^n) would reduce to the exact solution case of Eqs. (10) and (11), which means that ICD/GN reduces to ICD/NR in this case. In the present ICD/GN, we keep this approximation fixed for a whole iteration through the image before updating it. Powell[20] proved the global convergence of an algorithm of this type when the Hessian is positive definite everywhere in an open convex set. This corresponds geometrically to strict convexity as in our case, and we have observed stable convergence in all experiments.

Computation time is approximated by the multiplies and divides required to compute θ_1 and θ_2. Both in ICD/NR of Sec. 2.3, and in Eqs. (12) and (13), noting M_0 the number of nonzero projections, this results in $4M_0N$ operations per full image update with the appropriate storage of $y_i/(\hat{p}_i^k)^2$ and p^n. Therefore, ICD/GN and ICD/NR require approximately the same time for one iteration, including reevaluation of the expansion point. If we terminate these evaluations, θ_2 is constant. In addition, both are also equivalent in terms of the number of indexings through the projection matrix A, each requiring two, and the use of the approximation is not computationally more expensive.

The optimal scan pattern for the sequential pixel updates of the greedy ICD algorithm is not obvious, since nothing in the information carried by the sinogram dictates the best order in which the image pixels should be visited. The order of pixel updates affects convergence speed, and therefore may warrant study to determine the best method. Lexicographic scans which iterate in horizontal, then vertical coordinates every other iteration have advantages analytically,[24] but in our experiments here, simple repeated horizontal scans performed slightly better. Improved convergence speed also seems to be achieved using a random scan pattern.[25] Unless otherwise noted, all results presented here are from random patterns which visit each pixel once per scan. We present below comparisons in objective convergence between a simple lexicographic pattern and a random pattern.

3 Experimental Performances

Three different sets of data have been used to test the performance of MAP-type estimation with the ICD/GN algorithm. We first realized simulations with a synthetic head phantom in a 200×200 mm field with a total photon count $\approx 3.0 \times 10^6$. The original, with the FBP reconstruction, is shown in Fig. 1. In addition, real medical single photon emission computed tomography (SPECT) data of a human thorax from T99 sestamibi heart perfusion was used to illustrate a medical application. Reconstructions cover approximately 320×256 mm, with a total photon count $\approx 1.5 \times 10^5$ in this case. Finally, we considered a section of the Derenzo phantom from PET data to study the performance of the algorithm in a low SNR case ($\approx 8.8 \times 10^4$ counts). The Derenzo estimates were corrected for attenuation and detector sensitivity by means of a map of correction factors included in the Bayesian model. The attenua-
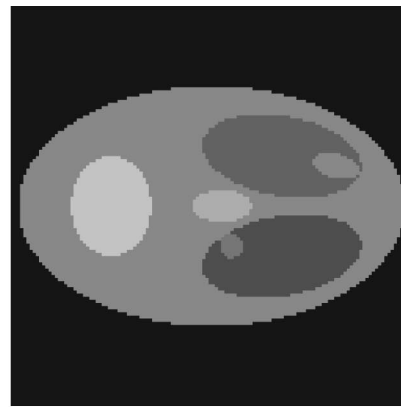

(a)
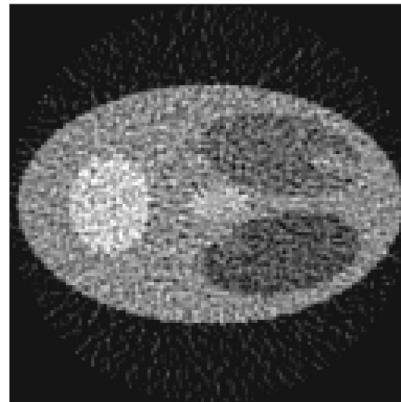
(b)

Fig. 1 Head phantom used for reconstruction comparisons. (a) Original phantom, (b) FBP reconstruction.

tion image itself was a MAP transmission image estimate for the phantom. The projection matrix was subsequently modified to incorporate the corrections as multiplicative factors. We also included a simple additive model to correct for random coincidences by adding an estimated bias of the emission counts. In addition to correcting photon statistics for random coincidences, this additive factor helps stabilize the convergence of the objective by moving the unbounded points of the likelihood function outside the constraint region. We chose an eight-point neighborhood system with normalized weights for the GGMRF, and investigated results for both Gaussian and non-Gaussian prior models. ML parameter estimation[26] provided the values of σ in the first two cases, while the last was chosen manually for best visual appearance. All reconstructions were initialized with the FBP image, and all estimates labeled ''exact MAP'' are computed via ICD/NR with random pixel scans for at least 100 iterations. Independently of the chosen number of updates of the global quadratic approximations, ICD/GN ran 30 and 40 iterations for the PET data, and the head and heart data, respectively.
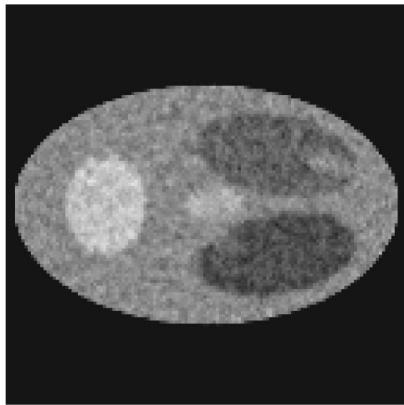
In each case, we have terminated the GN updates by fixing the quadratic approximation at a point we label $\bar{p}$. Following arrival at $\bar{p} = \hat{p}^k$, where $\hat{p}^k$ corresponds to the forward projection of the image estimate obtained after k iterations of the ICD/GN algorithm, the expansion point of
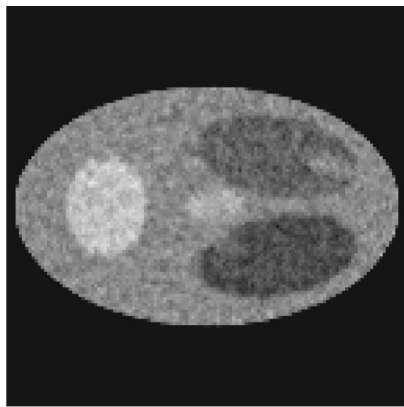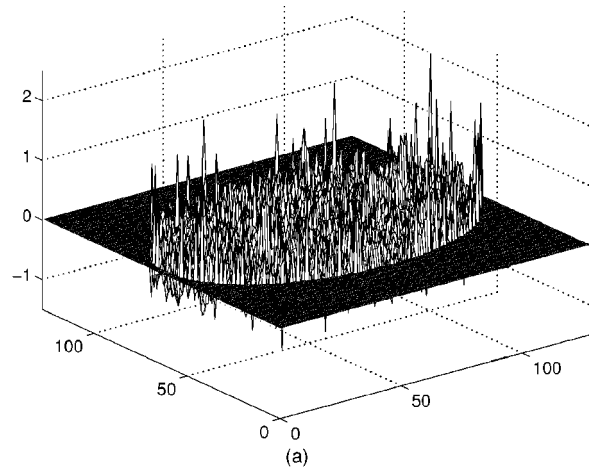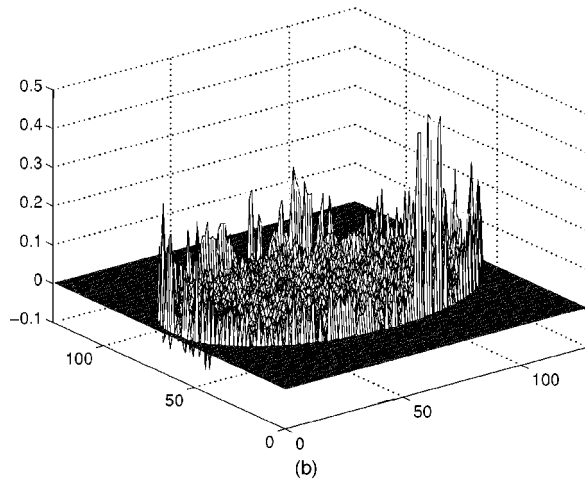
(a)
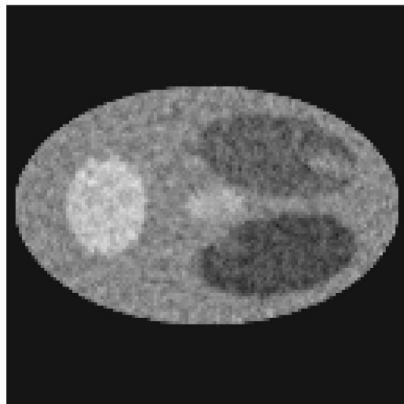
(b)

(c)

Fig. 2 Reconstructions of head phantom with Gaussian prior. (a) Exact MAP image, Gaussian prior and $\sigma = 0.58$; (b) Estimate with quadratic log-likelihood approximation fixed at $\bar{p} = y$; (c) ICD/GN result for quadratic approximation fixed after two iterations ($\bar{p} = \hat{p}^2 = Ax^2$).
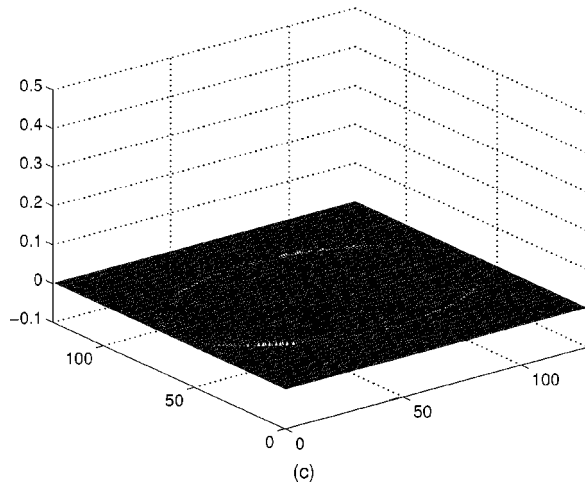

(a)

(b)

(c)

Fig. 3 MAP reconstruction with GGMRF, Gaussian prior ($q = 2.0$) and $\sigma = 0.58$. (a) Error between original phantom and MAP reconstruction; (b) difference between MAP and estimate using quadratic fixed at $\bar{p} = y$; (c) difference between MAP and estimate using quadratic fixed at $\bar{p} = \hat{p}^2 = Ax^2$.

the quadratic approximation is kept fixed for the remaining number of iterations. Referring to the two cases presented in Sec. 2.2, the parameters of the quadratic approximation may be computed directly from the measurement data ($\bar{p} = y$) and kept fixed, as in Eq. (5), or from the forward projection of the FBP image and updated at each iteration ($\bar{p} = \hat{p}^k = Ax^k$), as in Eq. (7), until the decision is made to stop the updates. In Figs. 2–7, we can see small but per-
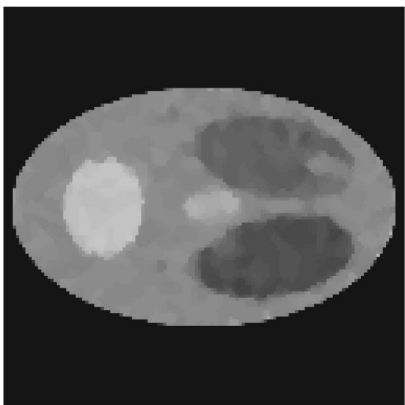
ceptible differences between the exact image and its approximation for $\bar{p} = y$ [(a), (b)], whereas those differences are not visible after updating the quadratic approximation only a small number of times [(a), (c)]. It is interesting to determine the smallest number of evaluations of the point
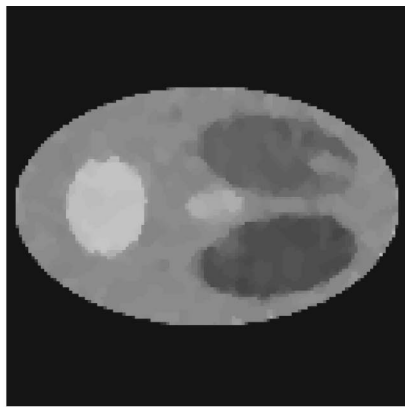
(a)
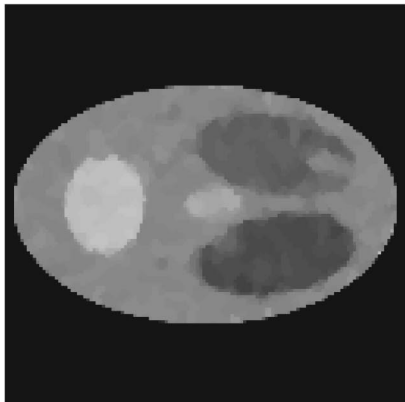
(b)

(c)

Fig. 4 Reconstructions of a head phantom with non-Gaussian GGMRF prior ($q=1.1$), $\sigma=0.25$. (a) Exact MAP image; (b) estimate with fixed quadratic approximation evaluated at $\bar{p}=y$; (c) ICD/GN result with fixed approximation after two iterations ($\bar{p}=\hat{p}^2=Ax^2$).

of expansion that is required to yield negligible difference between the images computed with ICD/NR and ICD/GN. Keeping the approximation fixed thereafter allows for solving the simple quadratic problem for the remaining iterations. A set of reconstructions with changing $\bar{p}=\hat{p}^k$ showed that after only two iterations (three evaluations of the expansion point, i.e., $\bar{p}=Ax^2$), the resulting final images show essentially no difference from the exact estimate [(a), (c), (e)]. For low SNR cases as in our SPECT and PET
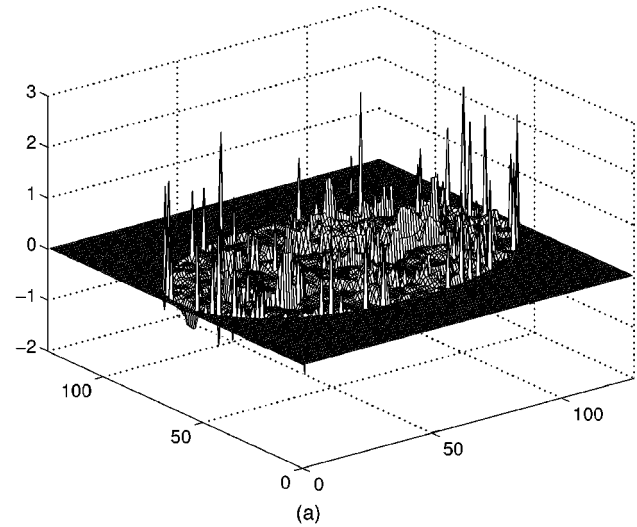
(a)
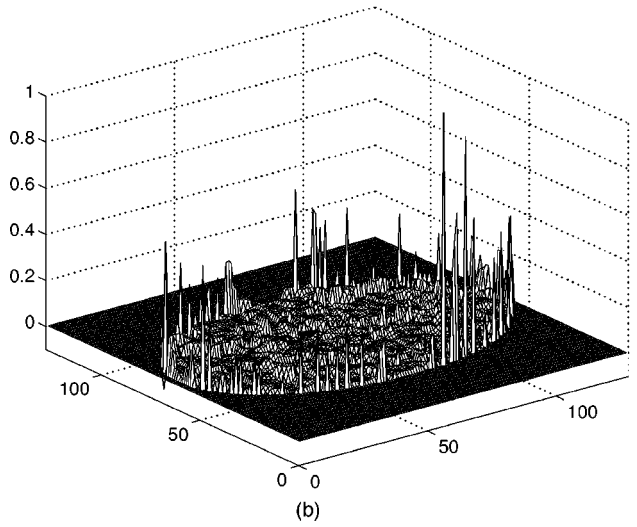
(b)
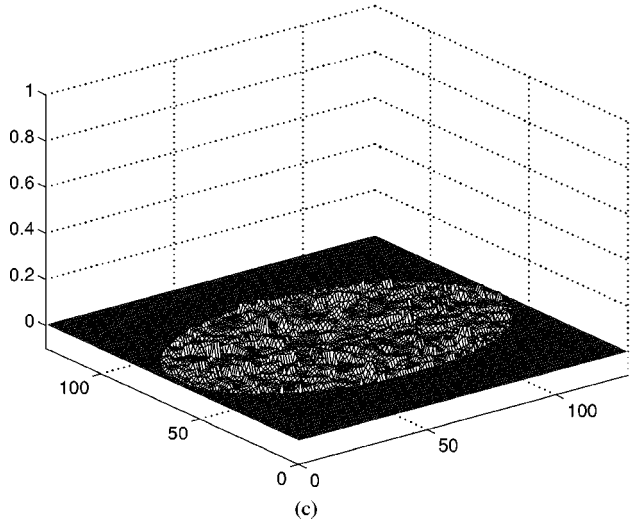
(c)

Fig. 5 MAP reconstruction with GGMRF, $q=1.1$, $\sigma=0.25$. (a) Error between original phantom and MAP reconstruction; (b) difference between MAP and estimate using quadratic at $\bar{p}=y$; (c) difference between MAP and estimate from quadratic using $\bar{p}=\hat{p}^2=Ax^2$.
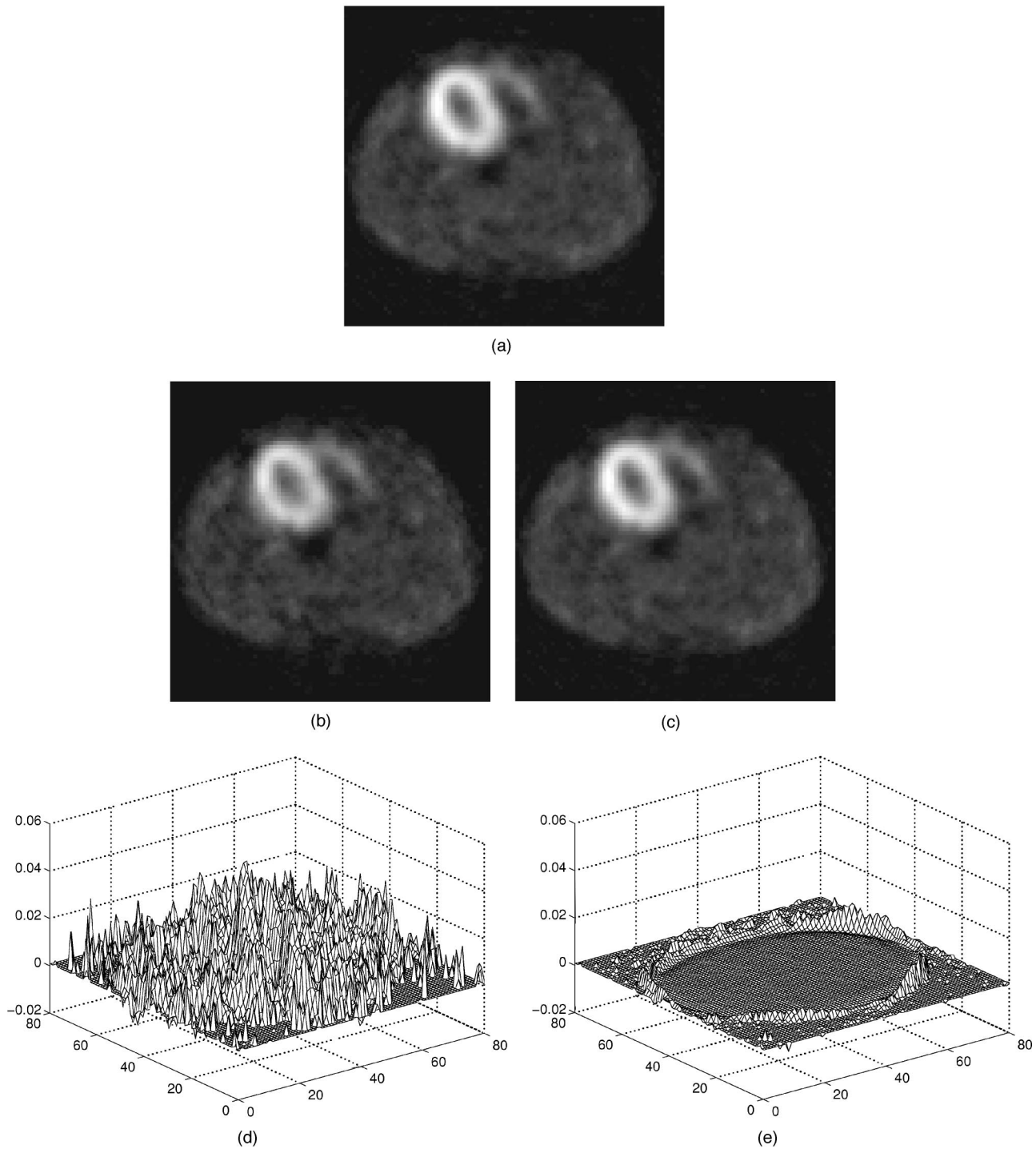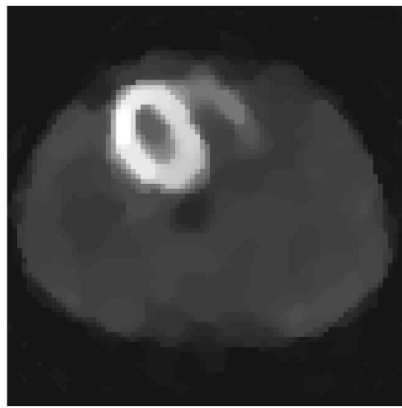
Fig. 6 Reconstructions of SPECT data with Gaussian prior and $\sigma = 0.028$. (a) Exact MAP image; (b) estimate using quadratic at $\bar{p} = y$; (c) ICD/GN result for two updates of the expansion point ($\bar{p} = \hat{p}^2 = Ax^2$); (d)(e) differences between (a) and (b),(c), respectively. Data courtesy of T. S. Pan and M. King, University of Massachusetts.
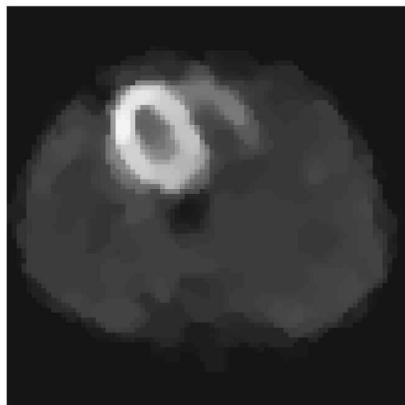
data, the algorithm uses a minimum pixel value exponentially decreasing to zero with iterations, in order to avoid frequently hitting the positivity constraint during the first iterations.

Table 1 shows error measures relative to the original image, to provide context for the magnitude of differences between the MAP and ICD/GN estimates of the head phan-
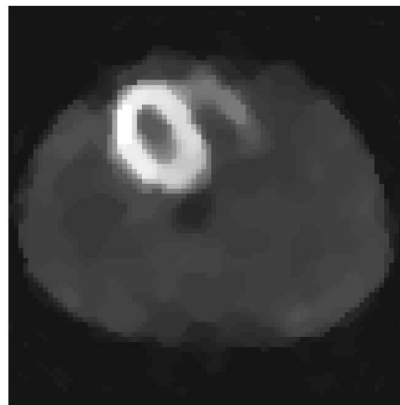
tom appearing in Table 2. The magnitudes of the differences displayed in Table 2 for different numbers of expansion updates show that the results are sufficiently accurate after one or two updates of the expansion first computed from the FBP image for ICD/GN to be very similar to an exact MAP reconstruction. It also indicates that with updating the expansion once per iteration during the complete
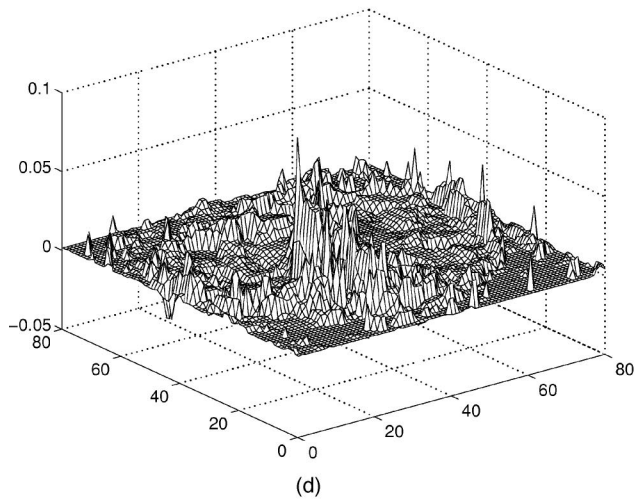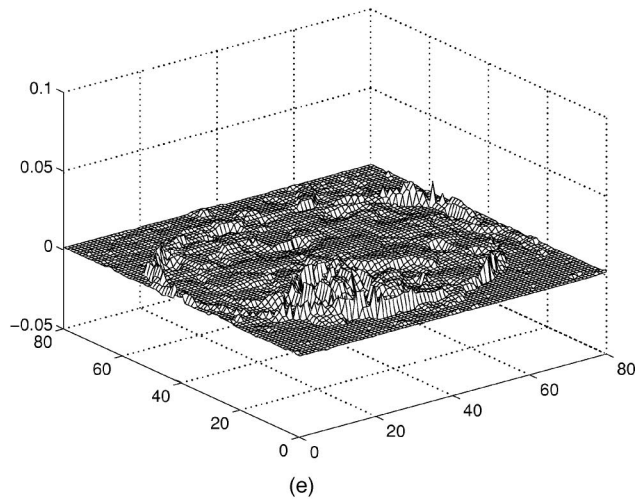
Fig. 7 Reconstructions of SPECT data with non-Gaussian GGMRF prior, $q=1.1$ and $\sigma=0.017$. (a) Exact MAP image; (b) estimate using quadratic at $\bar{p}=y$; (c) ICD/GN result for two updates of the expansion point ($\bar{p}=\hat{p}^2=Ax^2$); (d), (e) differences between (a) and (b), (c), respectively. Data courtesy of T. S. Pan and M. King, University of Massachusetts.

reconstruction process, the ICD/GN image appears to asymptotically converge to the exact estimate. Detailed study shows that the largest differences occur about the high frequency regions of the image, at the boundaries and outside the object, and especially in the regions where the photon counts are low. This relates to the studies in Refs. 15 and 8 showing that the larger the photon counts, the better the quadratic approximation to the log likelihood. This does not, however, visibly affect the quality of the reconstruction. In fact, differences computed only in the region of the object drop by a factor of 10 compared with the computation on the whole image. In the low SNR PET example of Fig. 8, we see similarly that after a total of three evaluations of the Taylor series for the log likelihood, we achieve a result nearly indistinguishable from the true MAP estimate computed by ICD/NR.

Table 1 Error in estimate relative to the original head phantom.

Reconstructed image	MSD	MAD	MAX
Head phantom	$q=2$	$\sigma=0.584$	
$\bar{p}=y$	0.0442	0.1076	2.2722
$\bar{p}=\hat{p}^0=Ax_{\text{FBP}}$	0.0456	0.1085	2.4000
$\bar{p}=\hat{p}^1=Ax^1$	0.0434	0.1073	2.2793
$\bar{p}=\hat{p}^2=Ax^2$	0.0428	0.1070	2.1711
$\bar{p}=\hat{p}^{40}=Ax^{40}$	0.0428	0.1070	2.1702
ICD/NR exact MAP estimate	0.0428	0.1070	2.1702

Interestingly, the quantitative results indicate that fixing the quadratic expansion from the forward projection of the FBP image is not necessarily better than using the raw sinogram with $\bar{p}=y$. The high frequency noise present in the original sinogram is spread over the entire image by the FBP reconstruction, and may therefore be spread over the entire sinogram resulting from the forward projection of the FBP image. In contrast, the vector y remains as the standard for the estimate's fidelity to the data throughout GN updates. Therefore, it is plausible that the minimum of the resulting GN quadratic approximation at $\bar{p}=\hat{p}^0=Ax_{\text{FBP}}$ may be farther from the MAP estimate than the one made at $\bar{p}=y$. The use of the FBP image as starting point for ICD/GN has several motivations. The FBP image is a good first estimate of image densities which is quickly computed. In addition, the ICD algorithm is very efficient in converging the high frequencies in the image, while somewhat slower for low frequencies; the FBP image offers low frequency components already near their optimal value. This

Table 2 Difference measurements between MAP and ICD/GN images for varying numbers of evaluations of the global quadratic approximation.

Expansion point	MSD	MAD	MAX
Head phantom	$q=2$	$\sigma=0.584$	
$\bar{p}=y$	4.663×10^{-4}	8.413×10^{-3}	5.160×10^{-1}
$\bar{p}=\hat{p}^0=Ax_{\text{FBP}}$	8.393×10^{-4}	6.00×10^{-3}	5.160×10^{-1}
$\bar{p}=\hat{p}^1=Ax^1$	1.325×10^{-4}	2.208×10^{-4}	2.327×10^{-1}
$\bar{p}=\hat{p}^2=Ax^2$	1.428×10^{-7}	6.685×10^{-5}	1.45×10^{-2}
SPECT data	$q=2$	$\sigma=0.0283$	
$\bar{p}=y$	3.969×10^{-5}	2.572×10^{-3}	5.88×10^{-2}
$\bar{p}=\hat{p}^0=Ax_{\text{FBP}}$	2.165×10^{-5}	1.813×10^{-3}	4.17×10^{-2}
$\bar{p}=\hat{p}^1=Ax^1$	3.772×10^{-6}	6.358×10^{-4}	2.43×10^{-2}
$\bar{p}=\hat{p}^2=Ax^2$	5.319×10^{-7}	2.165×10^{-4}	1.18×10^{-2}
Derenzo PET data	$q=2$	$\sigma=0.10$	
$\bar{p}=y$	4.468×10^{-4}	6.145×10^{-3}	2.736×10^{-1}
$\bar{p}=\hat{p}^0=Ax_{\text{FBP}}$	9.840×10^{-5}	3.105×10^{-3}	1.510×10^{-1}
$\bar{p}=\hat{p}^1=Ax^1$	2.490×10^{-5}	1.369×10^{-3}	1.024×10^{-1}
$\bar{p}=\hat{p}^2=Ax^2$	5.619×10^{-6}	4.919×10^{-4}	5.87×10^{-2}

makes the FBP image a good starting point for most iterative estimation algorithms. However, it is quite conceivable that in some instances of GN, one might still use $x^0=x_{\text{FBP}}$ with $\hat{p}^0=y$ in place of Ax_{FBP}.

The convergence plots of Fig. 9 correspond to the above results. They were obtained for Gaussian priors and used a regular scan pattern in horizontal coordinates first. The plots are similar in the case of non-Gaussian priors. For the head and the Derenzo phantoms, ICD/GN exhibits the same behavior as ICD/NR, indicating experimentally that the global quadratic approximations do not affect the convergence properties of the ICD algorithm. Whatever choice is made for fixing the quadratic approximation, the objective function converges quickly (typically about five iterations). However, it appears in the SPECT data plot that if we evaluate the quadratic approximation at $\bar{p}=y$ and keep it fixed thereafter ("0" plots), the limiting *a posteriori* probability is significantly different from the exact MAP objective. Updating the quadratic approximation only once after its first evaluation from the FBP image leaves only small differences in the objective values of ICD/GN and ICD/NR.

With the exception of the plots in Fig. 9, all the figures presented thus far result from ICD pixel updates in random order. The patterns analyzed in Ref. 24 were regular, by lines or by columns, but recent results have shown a potential for improvement in convergence speed with these randomly ordered scans.[25] Figure 10 presents the improvement in convergence speed that a random update pattern offers over a regular lexicographic scan in the case of the Derenzo phantom. A simple linear congruential random number generator guarantees that each pixel in the image is visited once and only once per iteration. The gain here is significant, since after only two iterations the objective has almost converged when the random scan is used, whereas six iterations are necessary when using the lexicographic scan. In addition, it is interesting to study the influence of the scan pattern on the ICD/GN results. Figures 10 and 11 compare the image estimates for $\bar{p}=\hat{p}^2$ when using the lexicographic or random update pattern. The magnitude plots show that the random pattern eliminates some of the differences between the GN estimate and the exact MAP image. With a sequential update of the pixels, the first pixels in the image tend to be overcorrected since the objective function is greedily optimized at each step. A lexicographic scan, such as in Figs. 10(b) and 11(a), therefore creates more low frequency artifacts in the image and tends to concentrate those pixels on whichever side of the object is updated first, whereas a random pattern scatters them. The differences in the above figures appear in fact to be the superposition of both the overcorrection discussed above, and the low frequency components which take longer to converge with ICD. Thus the random pattern may offer faster convergence of both the global approximations of GN, and the sweeps of the image under ICD. A potentially interesting variation on the random update would involve a nonuniform spatial distribution of updates dependent on the current image estimate.

These results suggest that any algorithm which converges rapidly for the first few iterations might quickly supply a final global quadratic likelihood approximation,
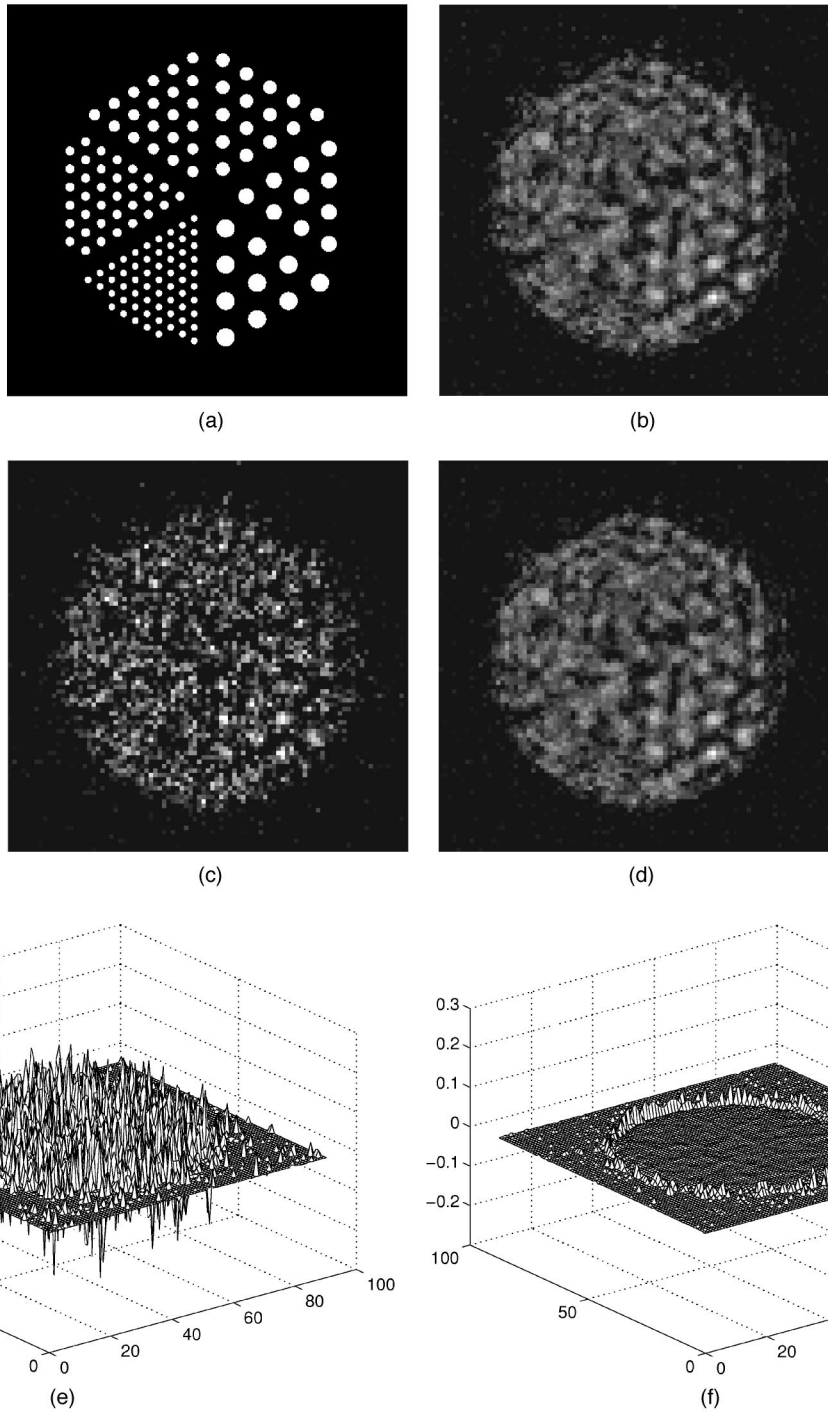
Fig. 8 Reconstructions of PET data with Gaussian prior. (a) Original Derenzo phantom; (b) MAP reconstruction with $\sigma=0.1$; (c) ICD/GN result for $\bar{p}=y$; (d) ICD/GN result for quadratic expansion fixed after two iterations ($\bar{p}=\hat{p}^2=Ax^{(2)}$); (e), (f) differences between MAP image of (b) and results in (c) and (d), respectively. Data courtesy of G. Hutchins, Indiana University.

which could then be attacked by any numerical algorithm well suited to quadratic objectives. An initially quite rapid, though nonconvergent, form of EM which cycles among subsets of data, known as ordered subsets EM (OSEM),[19] might be used for the initial estimates, and another, e.g., conjugate gradient or ICD, might be applied to convergence of the resulting fixed quadratic problem.[25]
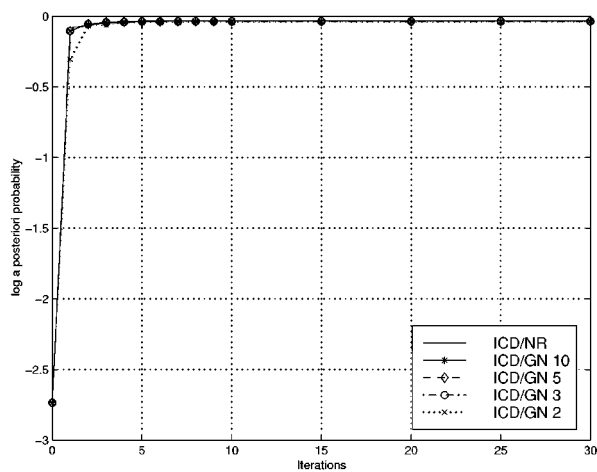
4 Conclusion

In Bayesian tomography, the substitution of an approximation to the log-likelihood function allows simpler optimization of the MAP objective without computation under the Poisson model. After only a few updates of the global quadratic, a coordinate descent method rapidly converges to an
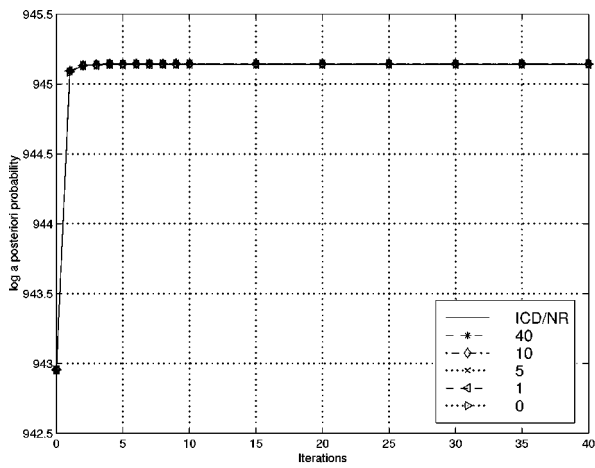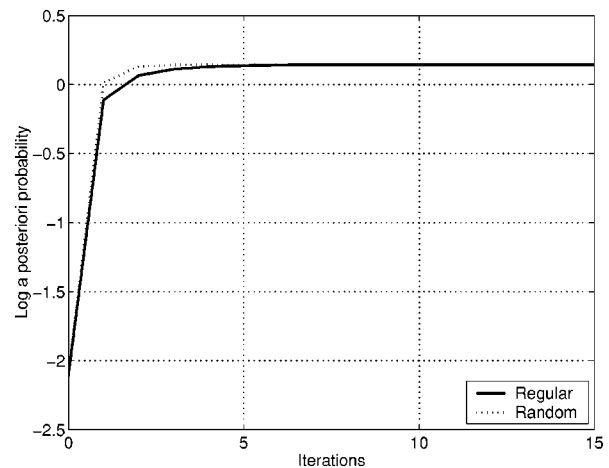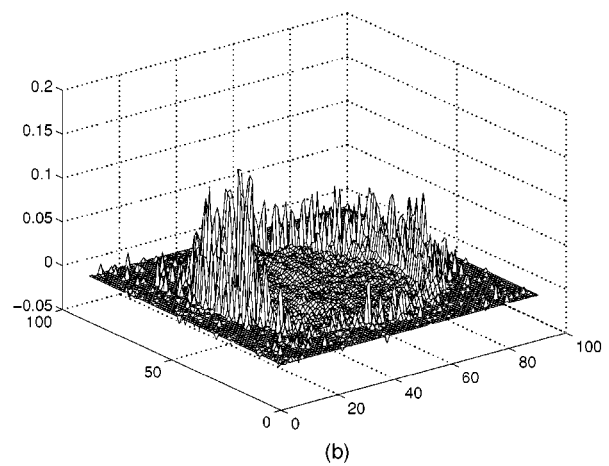
(a)
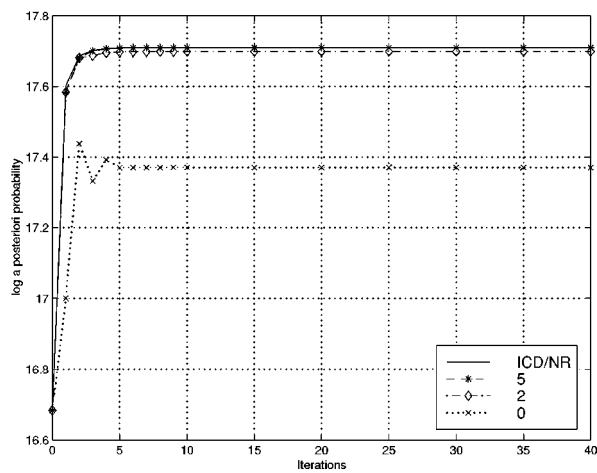
(b)

(c)

Fig. 9 Convergence of objective vs iterations for (a) the head phantom, (b) the SPECT data, and (c) the PET data, for various numbers of evaluations of the expansion point. Convergence plots obtained from ordered pixel updates in horizontal coordinates first in ICD. "0" indicates $\bar{p}=y$, "1" indicates $\bar{p}=\hat{p}^0=Ax_{\mathrm{FBP}}$, "2" indicates $\bar{p}=\hat{p}^1 =Ax^1$, etc.

(a)

(b)

(c)

Fig. 10 Difference in convergence behavior between lexicographic scan in horizontal coordinates first, and randomly ordered pixel updates in ICD, for the Derenzo phantom with Gaussian prior. (a) Convergence of log *a posteriori* probability in both cases. Residual difference between true MAP estimate and result from quadratic approximation fixed after two iterations ($\bar{p}=\hat{p}^2=Ax^2$) under (b) regular scan pattern; (c) random scan pattern.
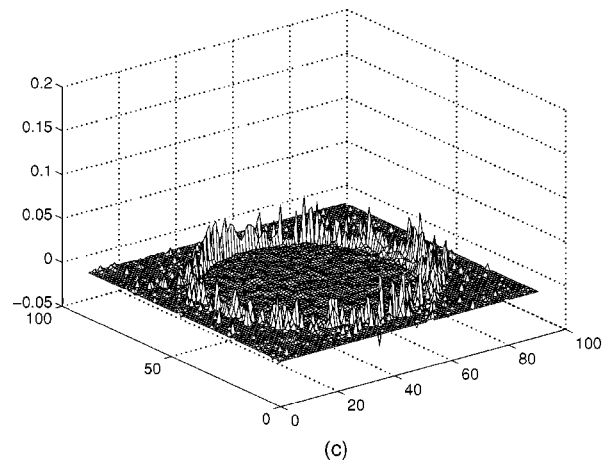
estimate nearly indistinguishable from the exact MAP estimate. For emission tomographic problems of medium to

high signal-to-noise ratios, our results suggest that viewing the log likelihood as quadratic may be adequate for any visual interpretations, and would open the problem to researchers and practitioners more accustomed to quadratic
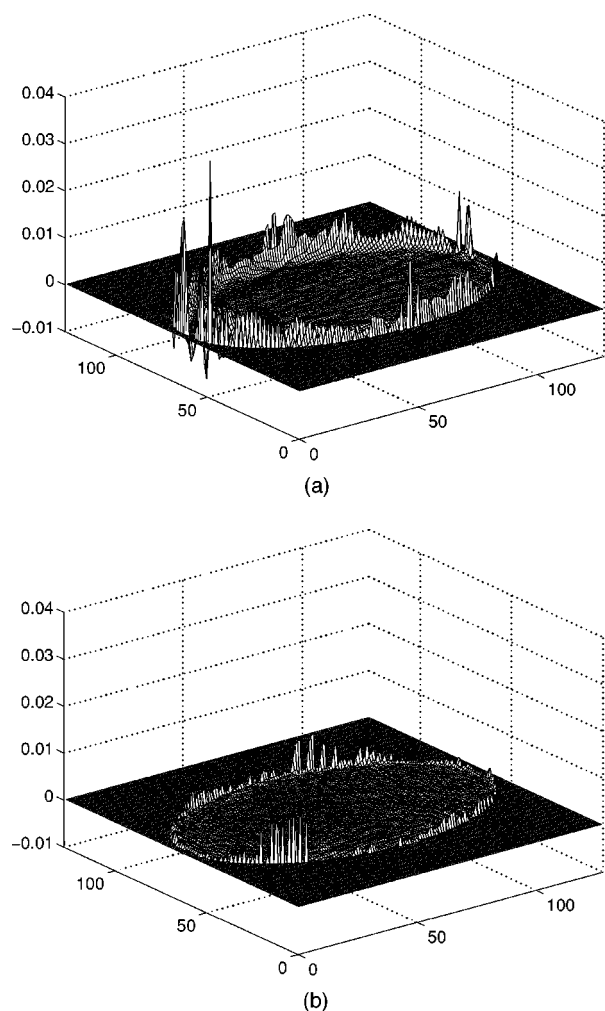
(a)

(b)

Fig. 11 Residual difference between true MAP estimate and ICD/GN result from quadratic approximation fixed after two iterations ($\bar{p} = \hat{p}^2 = Ax^2$) under (a) lexicographic scan pattern; (b) random scan pattern. (Head phantom with Gaussian prior, $q = 2.0$, $\sigma = 0.58$.)

constrained optimization than photon-limited image reconstruction. Even in low-count problems, it appears that little appreciable quality would be sacrificed in optimizing a quadratic log-likelihood approximation fixed very early in the process. These claims all assume the use of relatively rapidly converging techniques for the first few iterations. Further research will evaluate differences between ICD and OSEM for the initial stages.

Acknowledgment

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.*, pp. 12–18, Washington, DC (1985).

2. L. Shepp and Y. Vardi, ''Maximum likelihood reconstruction for emission tomography,'' *IEEE Trans. Med. Imaging* **MI-1**, 113–122 (1982).

3. R. Gordon and G. Herman, ''Three-dimensional reconstruction from projections: A review of algorithms,'' in *International Review of Cytology*, G. Bourne and J. Danielli, Eds., Vol. 38, pp. 111–151, Academic, New York (1974).

4. G. T. Herman and A. Lent, ''A computer implementation of a Bayesian analysis of image reconstruction,'' *Inf. Control.* **31**, 364–384 (1976).

5. E. Artzy, T. Elfving, and G. T. Herman, ''Quadratic optimization for image reconstruction,'' *Comput. Graph. Image Process.* **11**, 242–261 (1979).

6. S. L. Wood and M. Morf, ''A fast implementation of a minimum variance estimator for computerized tomography image reconstruction,'' *IEEE Trans. Biomed. Eng.* **BME-28**, 56–68 (1981).

7. B. Tsui, E. Frey, and G. Gullberg, ''Comparison between ML-EM and WLS-CG algorithms for SPECT image reconstruction,'' *IEEE Trans. Nucl. Sci.* **38**, 1766–1772 (1991).

8. J. Fessler, ''Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans,'' *IEEE Trans. Image Process.* **4**, 1439–1450 (1995).

9. P. Koulibaly, *Regularisation et Corrections Physiques en Tomographie d'Emission*, PhD thesis, University of Nice-Sophia Antipolis, France, Oct. 1996.

10. N. Clinthorne, T. Pan, P.-C. Chiao, W. Rogers, and J. Stamos, ''Preconditioning methods for improved convergence rates in iterative reconstruction,'' *IEEE Trans. Med. Imaging* **12**, 78–83 (1993).

11. 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. Imaging* **13**, 687–701 (1994).

12. D. Snyder and M. Miller, ''The use of sieves to stabilize images produced with the EM algorithm for emission tomography,'' *IEEE Trans. Nucl. Sci.* **NS-32**, 3864–3871 (1985).

13. J. Llacer and E. Veklerov, ''Feasible images and practical stopping rules for iterative algorithms in emission tomography,'' *IEEE Trans. Med. Imaging* **8**, 186–193 (1989).

14. C. A. Bouman and K. Sauer, ''A generalized Gaussian image model for edge-preserving map estimation,'' *IEEE Trans. Image Process.* **2**, 296–310 (1993).

15. C. A. Bouman and K. Sauer, ''A unified approach to statistical tomography using coordinate descent optimization,'' *IEEE Trans. Image Process.* **5**, 480–492 (1996).

16. L. Kaufman, ''Implementing and accelerating the EM algorithm for positron emission tomography,'' *IEEE Trans. Med. Imaging* **MI-6**(1), 37–51 (1987).

17. A. D. Pierro, ''A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography,'' *IEEE Trans. Med. Imaging* **14**(1), 132–137 (1995).

18. J. Fessler and A. Hero, ''Complete data spaces and generalized EM algorithms,'' in *Proc. IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, Vol. IV, pp. 1–4, Minneapolis, Minnesota (April 1993).

19. H. Hudson and R. Larkin, ''Accelerated image reconstruction using ordered subsets of projection data,'' *IEEE Trans. Med. Imaging* **13**, 601–609 (1994).

20. J. Dennis and R. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice–Hall, Englewood Cliffs, NJ (1983).

21. D. Luenberger, *Introduction to Linear and Nonlinear Programming*, Addison–Wesley, Reading, MA (1973).

22. S. Jacoby, J. Kowalik, and J. Pizzo, *Iterative Methods for Nonlinear Optimization Problems*, Prentice–Hall, Englewood Cliffs, NJ (1972).

23. S. Saquib, J. Zheng, C. A. Bouman, and K. D. Sauer, ''Provably convergent coordinate descent in statistical tomographic reconstruction,'' in *Proc. IEEE Int'l Conf. on Image Proc.*, Vol. II, pp. 741–744, Lausanne, Switzerland (September 1996).

24. K. Sauer and C. A. Bouman, ''A local update strategy for iterative reconstruction from projections,'' *IEEE Trans. Signal Process.* **41** (1993).

25. J. Bowsher, M. Smith, J. Peter, and R. Jaszczak, ''A comparison of OSEM and ICD for iterative reconstruction of SPECT brain images,'' *J. Nucl. Med.* **39**, 79P (1998).

26. C. A. Bouman and K. Sauer, ''Maximum likelihood scale estimation for a class of Markov random fields,'' in *Proc. IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, Vol. 5, pp. 537–540, Adelaide, South Australia (April 1994).

Charles A. Bouman received a B.S.E.E. degree from the University of Pennsylvania in 1981, and a MS degree in electrical engineering from the University of California at Berkeley in 1982. From 1982 to 1985, he was a staff member in the Analog Device Technology Group at the Massachusetts Institute of Technology, Lincoln Laboratory. In 1987 and 1989, he received MA and Ph.D. degrees in electrical engineering from Princeton University under the support of an IBM graduate fellowship. In 1989, he joined the faculty of Purdue University where he currently holds the position of Associate Professor in the School of Electrical and Computer Engineering. Professor Bouman's research interests include statistical image modeling and analysis, multiscale processing, and the display and printing of images. He is particularly interested in the applications of statistical signal processing techniques to problems such as fast image search and inspection, tomographic reconstruction, and document segmentation. Professor Bouman has 20 full journal publications, over 50 conference publications, and two awarded patents. He has performed research for numerous government and industrial organizations including National Science Foundation, U.S. Army, Hewlett-Packard, NEC Corporation, Apple Computer, Xerox, and Eastman Kodak. From 1991–1993, he was also an NEC Faculty Fellow. Professor Bouman is a member of IEEE, SPIE, and IS&T professional societies. He has been both chapter chair and vice chair of the IEEE Central Indiana Signal Processing Chapter, and an associate editor of the IEEE Transactions of Image Processing. He is currently a member of the IEEE Image and Multidimensional Signal Processing Technical Committee.

Ken Sauer was born in Decatur, Indiana. He received the BSEE in 1984 and the MSEE in 1985 from Purdue University, West Lafayette, IN, and the Ph.D. from Princeton University in 1989 as an AT&T Foundation fellow. He is currently an Associate Professor of Electrical Engineering at the University of Notre Dame. Professor Sauer is involved in research of statistical methods for tomographic image estimation and other nondestructive evaluation problems, numerical optimization and stochastic image modeling.

Jean-Baptiste Thibault was born in Rouen, France, in 1975. He graduated from Ecole Superieure d'Electricite, Gif-sur-Yvette, France, and completed the MS degree in electrical engineering from the University of Notre Dame, Indiana, in 1998. He is currently working in Global Technology Operations at General Electric Medical Systems. His research interests include numerical methods, Bayesian estimation, and applications to statistical image reconstruction.