ML Parameter Estimation for Markov Random Fields with Applications to Bayesian Tomography

Suhail S. Saquib, Charles A. Bouman, *Senior Member, IEEE,* and Ken Sauer, *Member, IEEE*

***Abstract*— Markov random fields (MRF's) have been widely used to model images in Bayesian frameworks for image reconstruction and restoration. Typically, these MRF models have parameters that allow the prior model to be adjusted for best performance. However, optimal estimation of these parameters (sometimes referred to as hyperparameters) is difficult in practice for two reasons: i) direct parameter estimation for MRF's is known to be mathematically and numerically challenging; ii) parameters can not be directly estimated because the true image cross section is unavailable.**

In this paper, we propose a computationally efficient scheme to address both these difficulties for a general class of MRF models, and we derive specific methods of parameter estimation for the MRF model known as generalized Gaussian MRF (GGMRF).

The first section of the paper derives methods of direct estimation of scale and shape parameters for a general continuously valued MRF. For the GGMRF case, we show that the ML estimate of the scale parameter, σ, has a simple closed-form solution, and we present an efficient scheme for computing the ML estimate of the shape parameter, p, by an off-line numerical computation of the dependence of the partition function on p.

The second section of the paper presents a fast algorithm for computing ML parameter estimates when the true image is unavailable. To do this, we use the expectation maximization (EM) algorithm. We develop a fast simulation method to replace the E-step, and a method to improve parameter estimates when the simulations are terminated prior to convergence.

Experimental results indicate that our fast algorithms substantially reduce computation and result in good scale estimates for real tomographic data sets.

I. Introduction

OVER THE past decade, Bayesian methods for image reconstruction and restoration have become increasingly popular because they allow accurate modeling of both data collection, and image behavior. For example, Bayesian methods have been widely studied for emission tomography [1]–[3], transmission tomography [4]–[6], and image restoration [7], [8].

While the model used for data collection (also known as the forward model) has varied depending on the application, most of these approaches have used Markov random fields (MRF's) to model the unknown image. This is because the MRF model is computationally tractable and can also capture many non-Gaussian aspects of images such as edges. A variety of continuously valued MRF models have been proposed for accurately modeling images [2], [3], [9], [10]–[13]. Most of these are distinguished by the choice of potential function that assigns cost to differences between neighboring pixels.

While Bayesian methods can improve the quality of reconstructed images, they also have the potential to create errors when the models do not accurately characterize the data. To avoid this problem, modern approaches usually include parameters that allow the prior model and/or forward model to be adjusted to achieve the best possible results for each data set. Often the prior model parameters are referred to as hyperparameters because their effect is only indirectly apparent through the measured data.

Ideally, model parameters must be estimated for each data set as part of the image reconstruction or restoration process. However, estimation of these model parameters is often difficult for two reasons. First, direct maximum likelihood (ML) estimation of MRF parameters from example images is known to be a difficult problem. This is because in most cases the normalizing constant of the distribution, known as the partition function, is an intractable function of the parameters.

Second, in most applications example images are not available for estimation of parameters. Instead, parameters must be estimated indirectly from collected data because the true image cross section is not known. This is a classic example of an incomplete data problem for which the expectation-maximization (EM) algorithm was developed [14], [15]. Intuitively, the EM algorithm works by iteratively computing the expectation of the unknown image statistics, and then maximizing the likelihood with respect to those statistics.[1]

Most previous research for the direct estimation of MRF parameters has focused on discrete MRF's [17]–[22] and used approximations to the ML estimate based on maximum pseudolikelihood [19], [20] or least squares [21]. The methods of these papers are not directly applicable to estimation of parameters from the continuously valued MRF's, which are the subject of this paper. More recently, researchers have developed methods that attempt to directly compute the partition function during the estimation process [23]–[26]. Since computation of the partition function requires on-line stochastic simulation, fast methods of simulation, such as those of [27] and [28], have been of considerable interest.

Manuscript received November 8, 1995; revised October 13, 1997. This work was supported by the National Science Foundation under Grant MIP93-00560. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Andrew Yagle.

S. S. Saquib is with Polaroid Corporation, Cambridge, MA 02139 USA (e-mail: saquibs@polaroid.com).

C. A. Bouman is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907 USA.

K. Sauer is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA.

Publisher Item Identifier S 1057-7149(98)04367-X.

[1] We note that this application of the EM algorithm is unrelated to Shepp and Vardi's classic use of EM in image reconstruction [16]. In Shepp and Vardi's work, the image is treated as the unknown parameter.

Alternatively, estimation of parameters for Gaussian MRF's has received wide attention. This problem is essentially equivalent to estimation of regularization parameters in quadratic regularization. For a review of these approaches, see [29]. We do not consider these methods since, for the purposes of this research, Gaussian prior models lead to excessive smoothing of image edges.

A number of researchers have specifically studied the problem of estimating continuous MRF parameters from incomplete data. The simplest and perhaps most natural approach to this problem is joint *maximum a posteriori* (MAP) estimation of both the image and parameters [30]. Unfortunately, this leads to an inconsistent estimator that may even be divergent [31], [32]. Mohammad-Djafari has reported good results by using a joint MAP estimation method which is stabilized with an appropriate prior distribution for the parameters [33]. Schultz, Stevenson, and Lumsdaine have proposed a method for ML estimation of parameters by employing a signal-dependent approximation to the partition function and a quadratic approximation to the posterior distribution of the image [34]. Recently, Higdon *et. al.* have independently proposed a method for sampling from the posterior distribution of the MRF parameters using direct precomputation of the partition function [35]. Samples from the posterior distribution can be used to compute confidence intervals, and if the posterior distribution is peaked, then individual samples are likely to be good estimates.

In an effort to avoid the computationally expensive E-step of the EM algorithm, Zhou and Leahy have developed an approach which uses a mean field theory approximation to compute the required expectation [32], [36]. This method is philosophically similar to mean field approximations used by Zhang in segmentation problems [37]. Pun and Jeffs have taken an approach similar in concept to EM, but replacing the ML step with a estimator designed specifically for the p parameter used in a generalized Gaussian MRF (GGMRF) [38], [39].

True EM approaches have often been avoided due to the perception of excessive computation. One of the earliest attempts was by Geman and McClure when they proposed computing the shape parameter of a MRF by precomputing the partition function and replacing the expectation step with stochastic integration [1], [40].

In this paper, we propose a computationally efficient scheme to compute ML estimates of MRF model parameters from incomplete observations [41]–[44]. Our method hinges on two innovations:

- simple direct parameter estimation for continuous MRF's based on closed form expressions for the dependence of the partition function on the scale parameter;
- computationally fast algorithms for computing the E-step of the EM algorithm based on fast simulation and parameter extrapolation.

In addition, we present a method for computing forward model parameters such as dosage which are often required for Bayesian reconstruction of transmission tomograms [45].

The first section of the paper derives methods for direct estimation of parameters for a general continuously valued MRF. To do this, we reparameterize many well known potential functions using two parameters referred to as scale and shape. We show that in the general case, the scale parameter, σ, may be computed as the numerical solution to a simple equation. Interestingly, for the specific case of the GGMRF [12], σ has a closed-form solution that is analogous to estimation of standard-deviation for Gaussian random variables [31]. Based on this result, we derive a general approach to ML estimation of the shape parameter through precomputation of a *one-dimensional (1-D) function* of the shape parameter. We then illustrate the method for the case of the GGMRF by showing computed values of σ and p for a variety of images.

The second section of the paper presents a fast algorithm, based on the EM algorithm, for computing ML parameter estimates when the true image is unavailable. We develop a fast simulation method to replace the E-step based on extensions to the conventional Metropolis algorithm [46]–[48], and the heuristics suggested by Green and Han [49]. For the case of both transmission and emission tomography problems, our updates may be efficiently computed using the techniques described in [50]. To further reduce computation, we introduce a method to extrapolate the parameter estimates when the simulations are terminated prematurely.

Experimental results are presented for real transmission and emission data sets, as well as for image restoration. These results indicate that our fast algorithms substantially reduce computational cost and result in useful scale parameter estimates.

II. ML Parameter Estimation for Continuously Valued MRF's

Let X be a continuously valued unknown random image, and let Y be the measured random observations. We use upper case letters to denote random quantities and lower case letters to denote their corresponding deterministic realizations. The unknown image is modeled by its probability density function, $\mathcal{P}_{\sigma,p}(x)$, where σ and p are unknown scale and shape parameters to be defined. We will assume that X contains N pixels indexed by X_s for $s \in S$, and that X takes values in the convex set $\Omega = \{x\colon x_s \geq 0 \text{ for all } s \in S\}$. The observations are modeled by their conditional density function, $\mathcal{P}(y|x)$. If Y is discrete, then $\mathcal{P}(y|x)$ is a probability mass function.

Our objective is to estimate the parameters σ and p so that we may compute the MAP estimate of X given Y

$$\hat{X}_{\text{MAP}} = \arg \max_{x \in \Omega} \{\log \mathcal{P}(Y|x) + \log \mathcal{P}_{\sigma,p}(x)\} \qquad (1)$$

where the constraint of $x \in \Omega$ enforces positivity in the image. In this section, we develop methods for computing ML parameter estimates from direct observations of X.

A. MRF Prior Models

We model the unknown image X as an MRF with Gibbs distribution

$$\mathcal{P}_{\sigma,p}(x) = \begin{cases} \dfrac{1}{\sigma^N z(p)} \exp\left\{-\dfrac{1}{p}\, u(x/\sigma, p)\right\}, & \text{if } x \in \Omega \\ 0, & \text{if } x \notin \Omega \end{cases} \qquad (2)$$

TABLE I
LIST OF NONCONVEX AND CONVEX POTENTIAL FUNCTIONS THAT HAVE BEEN USED

Nonconvex potential functions		
Author (name)	Reference	$\rho(\Delta, p)$
Geman and McClure	[1, 40]	$\frac{\Delta}{1+\Delta^2}$
Blake and Zisserman (weak spring)	[9, 51]	$\min\{\Delta^2, 1\}$
Hebert and Leahy	[2]	$\log(1+\Delta^2)$
Geman and Reynolds	[11]	$\frac{\lvert\Delta\rvert}{1+\lvert\Delta\rvert}$
Convex potential functions		
Author (name)	Reference	$\rho(\Delta, p)$
(Gaussian)		Δ^2
Besag (Laplacian)	[52]	$\lvert\Delta\rvert$
Green	[3]	$\log\cosh\Delta$
Stevenson and Delp (Huber)	[10]	$\min\{\lvert\Delta\rvert^2, 2\lvert\Delta\rvert - 1\}$
Bouman and Sauer (GGMRF)	[11]	$\lvert\Delta\rvert^p$

where σ is a parameter that controls scale or variation in X, and p is a shape parameter that we will see appears in many common MRF models. We use the notation p because this is the traditional variable used in the GGMRF model which will be the focus of our analysis. The function $u(x/\sigma, p)$ is referred to as the energy function.

The normalizing constant of the distribution, $\sigma^N z(p)$, is known as the partition function, and is computed as $\int_{x\in\Omega} \exp\{-(1/p)u(x/\sigma, p)\}\, dx$. It is easily verified that this function is proportional to σ^N where N is the number of pixels.

We consider energy functions of the form

$$u(x/\sigma, p) = \sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\rho\left(\frac{x_i - x_j}{\sigma}, p\right) \tag{3}$$

where $\mathcal{N}$ is the set of all neighboring pixel pairs, and $\rho(\cdot,\cdot)$ is the potential function which assigns a cost to differences between neighboring pixel values[2].

Depending on the choice of the potential function, (2) includes many common MRF models that have been proposed in the literature. Table I lists a variety of such potential functions. Notice that only the GGMRF model depends on p through the potential function. All other models listed depend on p solely through its explicit role in (2).

While these MRF models include Gaussian MRF's, Gaussian MRF's are of less interest because they tend to be excessively smooth. In an effort to better model image edges, a variety of functions have been suggested for $\rho(\Delta, p)$. These generally can be separated into functions which are convex or nonconvex in Δ. Convex potential functions often allow global optimization of (1) and are therefore computationally preferable. In addition, convex potential functions have been shown to lead to continuous or stable MAP estimates [12]. On the other hand, nonconvex potential functions tend to result in sharp discontinuities, which may be advantageous in applications such as edge detection.

[2]Notice that this distribution is not proper since its integral is infinite. The distribution may be made proper by adding terms of the form $\epsilon\rho(x_i/\sigma, p)$ to the energy function. All results of the paper may then be obtained by taking the limit as $\epsilon \to 0$.

TABLE II
LIST OF SCALABLE POTENTIAL FUNCTIONS. THE DIVERGENCE AND GENERALIZED DIVERGENCE ARE EXAMPLES OF SCALABLE POTENTIAL FUNCTIONS THAT REQUIRE TWO POSITIVE ARGUMENTS

Scalable potential functions			
Author (name)	Reference	$\rho(x_i, x_j, p)$	range of p
Bouman and Sauer (GGMRF)	[11]	$\lvert x_i - x_j\rvert^p$	$p > 0$
O'Sullivan (divergence)	[53, 13]	$(x_i - x_j)\log(x_i/x_j)$	p=1
(generalized divergence)	This paper	$(x_i - x_j)(x_i^{p-1} - x_j^{p-1})$	$p > 1$

The GGMRF model will be of particular interest to us because it will result in a simple closed-form expression for the ML estimate of σ. For this model, the density function for $X \in \Omega$ is given by

$$\mathcal{P}_{\sigma,p}(x) = \frac{1}{\sigma^N z(p)} \exp\left\{-\frac{1}{p\sigma^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i - x_j|^p\right\} \tag{4}$$

where normally $p \in [1, 2]$. Notice that (4) has a form that is analogous to a Gaussian distribution where σ plays the role of standard deviation. When $p = 2$, (4) reduces to a Gaussian model. Smaller values of p tend to produce sharper edges.

The GGMRF model has the advantage that its behavior is scale invariant [12], [13]. This property results from the fact that for all $x \in \Omega$, and $\sigma > 0$

$$u(x/\sigma, p) = \frac{1}{\sigma^p}\, u(x, p). \tag{5}$$

While Bouman and Sauer [12] showed that the function $|x|^p$ characterized all scale invariant functions of $\Delta = x_i - x_j$, Brette *et al.* [13] have shown that the class of scale invariant potential functions can be expanded if functions of both x_i and x_j are considered. For example, consider the divergence potential function proposed by O'Sullivan [53] and listed in Table II. This function also leads to the scalable property of (5) for $p = 1$, and therefore results in a scale invariant prior. In addition, the divergence function is known to be a convex function of (x_i, x_j) [54]. A third example, which we call the generalized divergence, is given in Table II. This function behaves like the divergence in the limit as $p \to 1$, but is well defined for $x_i = x_j = 0$. For a more detailed discussion of scale invariant priors, see [13].

B. ML Estimation of σ

We first derive the ML estimate of the scale parameter, σ, in the general case, and then specialize it to the GGMRF or any other scale invariant MRF model that obeys the property of (5).

The normalized log-likelihood may be computed from (2) to be

$$\frac{1}{N}\log\mathcal{P}_{\sigma,p}(x) = -\frac{1}{Np}\, u(x/\sigma, p) - \log\sigma - \frac{1}{N}\log z(p). \tag{6}$$

Differentiating (6) with respect to σ and equating the result to zero yields the equation for $\hat{\sigma}$, the ML estimate of σ [41].

$$\frac{\hat{\sigma}}{Np}\,\frac{\partial}{\partial\sigma}\,u(x/\sigma,p)\bigg|_{\sigma=\hat{\sigma}}=-1. \tag{7}$$

While this expression may look complex, $\hat{\sigma}$ may be easily evaluated to any desired precision using a standard root finding algorithm such as half interval search. This is interesting since in the general case ML estimation of MRF parameters is considered intractable due to the complex nature of the partition function. Note that a similar parameterization by Ogata and Tanemura [55] did not lead to such a simple solution due to the assumption that $\mathcal{P}(x)$ had bounded support.

For any scale invariant prior of Table II, we may evaluate the expression of (7) by substituting in the scaling relation of (5). This results in the simple expression

$$\hat{\sigma}^p=\frac{1}{N}\,u(x,p). \tag{8}$$

The above result is very appealing since it is quite simple, and applies for the GGMRF, divergence, and generalized divergence cases. In order to gain intuition, consider the case when X_i are i.i.d. Gaussian random variables. In this case $p=2,\sigma^2$ is simply the variance, and (8) reduces to the familiar expression $\hat{\sigma}^2=(1/N)\,\Sigma_{i=1}^{N}\,x_i^2$. Lange obtained a result equivalent to (8) in [31].

C. Joint ML Estimate of σ and p for Scalable Priors

In this section, we will derive a method for computing the joint ML estimate of p and σ for the GGMRF model or any other model that obeys (5).

We can reduce this problem to a 1-D optimization since we have the closed-form ML estimate of σ (8) in terms of p. To do this, we substitute the ML estimate of σ from (8) into the log-likelihood function of (6).

$$\frac{1}{N}\,\log\mathcal{P}_{\hat{\sigma},p}(x)=\frac{-1}{p}-\frac{\log(u(x,p)/N)}{p}-\frac{\log z(p)}{N} \tag{9}$$

The ML estimate of p is then given by

$$\hat{p}=\arg\,\min_{p}\left\{\frac{\log(u(x,p)/N)}{p}+\frac{1}{p}+\frac{\log z(p)}{N}\right\}. \tag{10}$$

In this form, we can see that the function $\sigma^p=u(x,p)/N$ is a sufficient statistic for the parameter p.

The first two terms of (10) are easily computed, but the third term, $\log\ z(p)/N$, is generally an intractable function of p since it requires the computation of a N-dimensional integral. However, we may compute the partition function indirectly through its derivative using stochastic simulation [56]. The derivative is given by

$$\begin{aligned}\frac{d}{dp}\,\frac{\log z(p)}{N}&=\frac{d}{dp}\,\frac{1}{N}\,\log\int_{x\in\Omega}\exp\left\{-\frac{1}{p}\,u(x,p)\right\}dx\\&=\frac{-1}{Nz(p)}\int_{x\in\Omega}\left(\frac{d}{dp}\,\frac{1}{p}\,u(x,p)\right)\\&\quad\cdot\exp\left\{-\frac{1}{p}u(x,p)\right\}dx\\&=\frac{1}{p^2}-\frac{1}{Np}\,E\left[\frac{d}{dp}\,u(X,p)\bigg|\sigma=1,p\right]\end{aligned} \tag{11}$$

where the last equality uses the consistency of the ML estimator for σ. The rest of the development in this section will be for the GGMRF prior. The extension to other scalable priors is similar.

Rewriting (11) for the GGMRF prior, we obtain

$$\begin{aligned}\frac{d}{dp}\,\frac{\log z(p)}{N}=\frac{1}{p^2}-\frac{1}{Np^2}\sum_{\{i,j\}\in\mathcal{N}}b_{i-j}\,E[|X_i-X_j|^p\\\cdot\log\left(|X_i-X_j|^p\right)|\sigma=1,p]\end{aligned}$$

where the function $\Delta^p\log\left(\Delta^p\right)$ is interpreted to be zero for $\Delta=0$. Next, define the function $f(p)$ so that

$$\begin{aligned}\frac{df(p)}{dp}=\frac{-1}{Np^2}\sum_{\{i,j\}\in\mathcal{N}}b_{i-j}\,E[|X_i-X_j|^p\\\cdot\log\left(|X_i-X_j|^p\right)|\sigma=1,p].\end{aligned} \tag{12}$$

Then the ML estimate of p is given by

$$\hat{p}=\arg\min_{p}\left\{\frac{\log(u(x,p)/N)}{p}+f(p)\right\}. \tag{13}$$

The minimization of (13) may be evaluated by first computing $f'(p)$ of (12) using stochastic integration. The stochastic integration may be done by generating samples from the desired MRF using $\sigma=1$, and computing the desired average. We note that this result rests on the reduction of the two-dimensional (2-D) parameter estimation problem to a 1-D problem, since in 1-D a derivative of $f'(p)$ is easily integrated to yield $f(p)$.

Note that all expectations are normalized by N. While the $\lim_{N\to\infty}\,\log\,\mathcal{P}_{\sigma,p}(x)$ generally does not exist, the normalized log-likelihood, $\lim_{N\to\infty}\,(1/N)\,\log\,\mathcal{P}_{\sigma,p}(x)$ does. Therefore, if we compute $f(p)$ for a sufficiently large lattice, we may assume that it does not vary with N.

D. ML Estimate of σ and p for Nonscalable Priors

In this section, we derive methods to compute the joint ML estimates of σ and p when the potential function is not scalable. This includes all the potential functions of Table I except the Gaussian, Laplacian, and GGMRF.

Notice that $u(x,p)$ is not a function of p for any of the nonscalable potential functions. This means that for $\sigma=\hat{\sigma}$ the log-likelihood of (6) may be simplified to be

$$\frac{1}{N}\,\log\,\mathcal{P}_{\hat{\sigma},p}(x)=\frac{-1}{Np}\,u(x/\hat{\sigma})-\log\,\hat{\sigma}-\frac{\log\,z(p)}{N}. \tag{14}$$

where $\hat{\sigma}$ is given by (7). The term $\log\ z(p)/N$ may be computed in a manner similar to $f(p)$ by first computing its derivative.

$$\frac{d}{dp}\,\frac{\log z(p)}{N}=\frac{1}{Np^2}\,E[u(X)\mid\sigma=1,p]. \tag{15}$$

Therefore the solution may be computed as the simultaneous solution to (7) and

$$\hat{p}=\arg\,\min_{p}\left\{\frac{u(x/\hat{\sigma})}{Np}+\log\,\hat{\sigma}+\frac{\log z(p)}{N}\right\}.$$

III. Parameter Estimation from Incomplete Data

The previous section dealt with the problem of estimating the prior model parameters σ and p from the observed image X. However, in many applications the image X is never directly observed. For example, in tomography the photon counts, Y, are only indirectly related to the image X. In this case, there may be additional parameters, ϕ, related to the forward model, $\mathcal{P}_\phi(y|x)$.

Ideally, the ML parameter estimate is then given by

$$(\phi, \sigma, p) = \arg \max_{(\phi,\sigma,p)} \int_{x\in\Omega} \mathcal{P}_\phi(y|x)\mathcal{P}_{\sigma,p}(x)\, dx. \tag{16}$$

While (16) is often difficult to compute directly, the EM algorithm is an effective method for iteratively maximizing (16) [14], [15].

In order to simplify notation, we will use the parameterization (γ, p) where $\gamma = \sigma^p$. Then a single update of the EM algorithm is given by

$$\phi_{k+1} = \arg \max_{\phi} E[\log\ \mathcal{P}_\phi\ (y|X)|Y = y, \phi_k, \gamma_k, p_k] \tag{17}$$

$$(\gamma_{k+1}, p_{k+1}) = \arg \max_{(\gamma,p)} E[\log\ \mathcal{P}_{\gamma,p}(X)|Y = y, \phi_k, \gamma_k, p_k] \tag{18}$$

where γ_k and p_k are the parameters generated at the kth iteration of the EM algorithm. It can be shown that each iteration of the EM algorithm increases the likelihood, so that the likelihood value is guaranteed to converge to a local maximum.

For the GGMRF prior, the EM update of (18), may be explicitly computed as

$$p_{k+1} = \arg \min_{p} \cdot \left\{ \frac{1}{p}\ \log\ E[u(X,p)/N|Y = y, \phi_k, \gamma_k, p_k] + f(p) \right\} \tag{19}$$

$$\gamma_{k+1} = \frac{1}{N}\ E[u(X, p_{k+1})|Y = y, \phi_k, \gamma_k, p_k]. \tag{20}$$

The expectations of (19) and (20) may be approximated using *on-line* stochastic integration. This is done by generating samples from the posterior distribution of X given Y, and then computing the desired sample averages in place of the true expectations [1], [22], [40]. Unlike the off-line stochastic integration of (12), evaluation of EM updates must be computationally efficient. In Section III-C, we will show how this is possible.

If p is known, then only σ needs to be estimated. In fact, estimation of σ is essential in many inverse problems. Too small a value of σ results in overly smooth images and too large a value of σ results in images with excessive noise. For this case, only (20) need be applied, and the expectation may be computed by averaging values of $u(X,p)$ for multiple samples of the image X generated from the posterior distribution of X given Y. We will discuss efficient algorithms for generating these samples in the next sections.

If both p and σ must be estimated, then the update of (19) must be computed first, and the result used to compute (20). Computation of (19) is somewhat more difficult since it requires that multiple samples of X be stored so the expectation may be computed as a function of p. However, we will show that often a single sample of X is sufficient to perform each EM update, so only a single image need be stored.

For nonscalable priors, the new parameters σ_{k+1} and p_{k+1} are given by the solution to the coupled equations.

$$p_{k+1} = \arg \min_{p} \left\{ E\left[\frac{u(X/\sigma_{k+1})}{Np} \middle| Y = y, \phi_k, \sigma_k, p_k \right] + \log\ \sigma_{k+1} + \frac{\log z(p)}{N} \right\} \tag{21}$$

$$\frac{\sigma_{k+1}}{Np_{k+1}}\ E\left[\frac{d}{d\sigma}\ u(X/\sigma)\bigg|_{\sigma=\sigma_{k+1}} |Y = y, \phi_k, \sigma_k, p_k \right] = -1. \tag{22}$$

These equations may be solved by iteratively computing the solution to each. Since each equation represents the minimization with respect to the corresponding variable, iterative solution will not diverge (if the ML estimate exists). When p is assumed known, the EM update for σ is given by (22) alone. However, computing the expectation of (22) requires buffering of the sample images.

A. Stochastic Data Models for Tomography

In this section, we introduce the stochastic models that we will need for emission and transmission tomography. For a description of photon counting models in tomography see [16] and [4], and for a development which is notationally similar to the one presented here, see [50].

Let x denote the column vector of emission intensities in the emission case or the attenuation densities in the transmission case. For the emission case, let A_{ij} be the probability that a photon emitted from cell j is registered at the ith detector. Let A be the projection matrix with elements $\{A_{ij}\}$, and let A_{i*} denote the ith row of the projection matrix. Let y denote the column vector of measurements of Poisson-distributed photon counts at the detectors for all angles and displacements. Then, for the emission case, the log conditional distribution of the photon counts Y given x is

$$\text{(emission)}\ \log\ \mathcal{P}(Y = y|x) = \sum_i \left(-A_{i*}x + y_i\ \log\ \{A_{i*}x\} - \log(y_i!)\right). \tag{23}$$

This formulation is general enough to include a wide variety of photon-limited imaging problems, and the entries of A may also incorporate the effects of detector response and attenuation.

The transmission case is similar, but with A_{ij} corresponding to the length of intersection between the jth cell and the ith projection. Let the input photon counts be Poisson-distributed

with rate y_T. Then the conditional log-likelihood of Y given x for the transmission case is

$$\text{(transmission) } \log\ \mathcal{P}(Y=y|x) = \sum_i \left(-y_T e^{-A_{i*}x} + y_i(\log\ y_T - A_{i*}x) - \log(y_i!)\right). \tag{24}$$

B. ML Estimate of Dosage y_T

The data for transmission tomography is often recorded in the form $z_i = \ln(y_T/y_i)$. This preserves the ML estimates of integral densities, but results in the loss of the parameter y_T which is required for the log-likelihood of (24). If y_T is unknown, it can be estimated along with other parameters using the EM algorithm update equation of (17).

Using the result of the appendix, we obtain the following EM update equation for y_T

$$y_{T_{k+1}} = \frac{M}{2}\left\{\sum_{i=1}^{M} E[\{e^{-A_{i*}X} + A_{i*}Xe^{-z_i}\} |Z=z,\sigma_k,p_k,y_{T_k}] - e^{-z_i} - z_i e^{-z_i}\right\}^{-1}. \tag{25}$$

C. Fast Simulation Technique

The EM parameter updates derived in Section III require the expectation of functions of X. Direct computation of these expectations is intractable, but we can approximate them by first generating sample images from the posterior distribution of X given Y and then computing averages using the sample images. The well-known Metropolis algorithm [46] can be used to generate these samples from the posterior distribution, but it tends to suffer from slow convergence.

In this section, we propose a faster simulation method based on the algorithms of Hastings [47] and Peskun [48]. The experimental results indicate that the required expectations can be accurately estimated using only a single image sample.

Let $q(x'|x)$ be an arbitrary transition probability for generating a new state x' from the current state x. Then in order to generate a sample with distribution $\pi(x)$, one should accept new samples with probability

$$\alpha(x',x) = \min\left\{1,\ \frac{q(x|x')\pi(x')}{q(x'|x)\pi(x)}\right\}.$$

The Metropolis algorithm is a special case of this general formulation when we choose $q(x'|x) = q(x|x')$. Another special case is the Gibbs sampler [8] when the new state for pixel j is generated using the conditional distribution, under $\pi(x)$, of x_j given the values of all other pixels. For the Gibbs sampler, $\alpha(x,x') = 1$, and we always accept the new state.

A good choice of transition probability $q(x'|x)$ results in faster convergence of the stochastic simulation; but at present, the optimal selection of $q(x'|x)$ is an open problem. For the tomography problem, it has been shown that greedy pixel-wise optimization of the posterior distribution has fast convergence [6], [50]. We therefore conjecture that the Gibbs sampler is desirable for the tomography problem because each new pixel is generated from its marginal posterior distribution. Experimental results presented in Section IV-B will support this conjecture.

Let us first examine the form of the conditional distribution of x_j required by the Gibbs sampler. Let x^n be the image at the nth iteration. Then for the emission case, from (23), (2) and (3), we have

$$\log\ \mathcal{P}(x_j|\{X_k = x_k^n\colon k\neq j\},y) = \sum_i \left(-A_{ij}x_j + y_i\ \log\{A_{ij}(x_j - x_j^n) + A_{i*}x^n\}\right) - \frac{1}{p}\sum_{k\in\mathcal{N}_j} b_{j-k}\,\rho\left(\frac{x_j - x_k^n}{\sigma}, p\right) + C \tag{26}$$

where C is constant independent of x_j and $x_j > 0$. Note that directly generating samples from (26) would be very computationally expensive. Green and Han [49] suggested using a Gaussian distribution instead with parameters chosen to approximate the transition distribution of the Gibbs sampler. However, due to the non-Gaussian nature of our prior term, this approximation is good only for the data term [50] in (26). We can therefore obtain a good approximation by retaining the prior term as it is and using a second order Taylor series expansion for the data term of (26)

$$\log\ \mathcal{P}(x_j|\{X_k = x_k^n\colon k\neq j\},y) \approx d_1(x_j - x_j^n) + \frac{d_2}{2}(x_j - x_j^n)^2 - \frac{1}{p}\sum_{k\in\mathcal{N}_j} b_{j-k}\,\rho\left(\frac{x_j - x_k^n}{\sigma}, p\right) + C' \tag{27}$$

where d_1 and d_2 are the first and second derivative of the data term with respect to x_j evaluated at x_j^n. In [50], it is shown that for the emission case

$$d_1 = -\sum_i A_{ij}\left(1 - \frac{y_i}{\tilde{p}_i^n}\right)$$

$$d_2 = -\sum_i y_i\left(\frac{A_{ij}}{\tilde{p}_i^n}\right)^2$$

where $\tilde{p}^n = Ax^n$. The approximation holds for the transmission case also, with the corresponding expressions for d_1 and d_2 as follows:

$$d_1 = -\sum_i A_{ij}\left(y_i - y_T e^{-\tilde{p}_i^n}\right)$$

$$d_2 = -\sum_i A_{ij}^2 y_T e^{-\tilde{p}_i^n}.$$

For efficient computation, we keep $\tilde{p}^n$ as a state vector and update it after each pixel update as follows:

$$\tilde{p}^{n+1} = A_{*j}(x_j^{n+1} - x_j^n) + \tilde{p}^n.$$

Let the transition distribution for generating the new state for the jth pixel be denoted as $q_j(x)$. Then we would like $q_j(x)$ to be a Gaussian distribution with mode m equal to the mode of the approximated conditional distribution (27). Unfortunately, generating positive samples from $q_j(x)$ is computationally intensive when $m \ll 0$. However, we

can use the fact that the tail of a Gaussian distribution may be accurately approximated as an exponential distribution. In the light of the above discussion, we choose $q_j(x)$ with the following form:

$$q_j(x) = \begin{cases} \dfrac{1}{C(s,m)} \exp\left\{\dfrac{-(x-m)^2}{2s^2}\right\} & m>0, x>0 \\ \dfrac{1}{\beta}\exp\left\{\dfrac{-x}{\beta}\right\} & m<0, x>0 \\ 0 & x<0 \end{cases}$$

where $C(s,m)$ is the normalizing constant of the truncated Gaussian distribution and m is the mode of the approximated conditional distribution (27)

$$m = \arg\max_x \left[d_1(x-x_j^n) + \frac{d_2}{2}(x-x_j^n)^2 - \frac{1}{p}\sum_{k\in\mathcal{N}_j} b_{j-k}\,\rho\left(\frac{x-x_k^n}{\sigma}, p\right)\right]. \quad (28)$$

Choosing s^2 is more difficult due to the prior term. Since we can at best do an approximate fit to the original distribution (26), it is not clear whether a more precise choice of s^2 would yield a significant improvement in performance. We therefore choose

$$s^2 = \frac{1}{d_2}$$

to be the variance of the data term. Note that the variance of the approximated distribution (27) is over estimated by this particular choice. β is determined by setting the derivative of the log of the exponential distribution equal to the derivative of (27) at $x_j = 0$

$$\beta = \left\{ -d_1 + d_2 x_j^n + \frac{1}{\sigma p}\sum_{k\in\mathcal{N}_j} b_{j-k}\,\rho'\left(\frac{-x_k^n}{\sigma}, p\right)\right\}^{-1}$$

where $\rho'(\cdot,\cdot)$ is the derivative of $\rho(\cdot,\cdot)$ with respect to its first argument.

Once d_1 and d_2 are computed, the optimization of (28) is computationally inexpensive since the sum associated with the prior typically involves few pixels. We use the half interval method to compute m. Note that during MAP reconstruction, m is also the updated value of pixel x_j in the iterative coordinate descent (ICD) algorithm of Bouman and Sauer [50].

D. Extrapolation of Parameter Estimates

Even with exact computation of the E-step, the convergence of the EM algorithm can sometimes be slow. One way to further reduce the computation is to improve the current EM parameter estimates by extrapolating them. This extrapolation requires very little computation, so it may be applied at each EM iteration. At each iteration k, the extrapolated parameter, $\theta_k^{(e)}$, is then an improvement over the EM parameter θ_k.

First consider the case of the GGMRF where p is known and $\gamma = \sigma^p$ must be estimated. This is an important special case. It is well known that

$$\frac{d}{d\gamma} E[\log \mathcal{P}_\gamma(X)|Y=y,\hat{\gamma}]\Big|_{\gamma=\hat{\gamma}} = 0$$

where $\hat{\gamma}$ is the ML estimate of γ. From this it can be shown that

$$\hat{\gamma} = E[u(X,p)/N|Y=y,\hat{\gamma}]. \quad (29)$$

The EM algorithm iteratively solves for the fixed point of this equation. However, a faster method is to search directly for its root. Define the function

$$g(\gamma) = E[u(X,p)/N|Y=y,\gamma] - \gamma. \quad (30)$$

Then the ML estimate of γ is the solution to $g(\hat{\gamma}) = 0$. At iteration k of the EM algorithm, the value of $g(\gamma_k)$ is given by

$$\begin{aligned} g(\gamma_k) &= E[u(X,p)/N|Y=y,\gamma_k] - \gamma_k \\ &= \gamma_{k+1} - \gamma_k. \end{aligned} \quad (31)$$

Therefore, we can plot the function $g(\gamma)$ computed from (31).

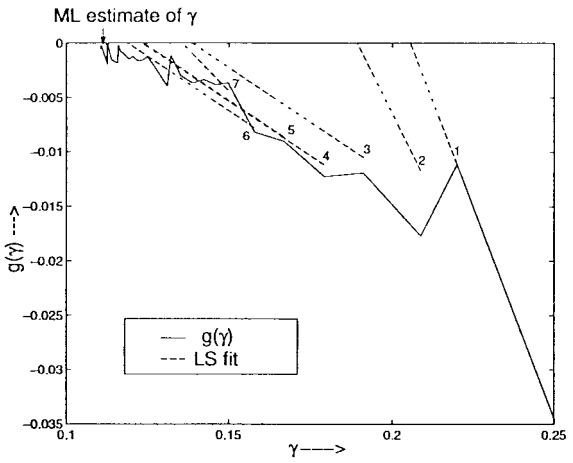


Fig. 1. The plot shows the normalized gradient $g(\gamma)$ computed at the EM updates of γ_k for a emission phantom using a GGMRF prior with $p = 1.1$. The ML estimate of γ is given by the root of $g(\gamma)$. The least squares (LS) fit obtained at the first seven points are numbered (1–7) and shown with a dashed line. The intersection of the dashed lines with the top of the graph are the extrapolated parameter values, $\gamma_k^{(e)}$.

Fig. 1 shows such a plot of $g(\gamma)$ with respect to γ for an emission phantom when we use only one sample of X to estimate the expectation of $u(X,p)$. At each iteration, we use the last three values of $(\gamma_k, g(\gamma_k))$ to compute a least squares fit to a straight line. The zero crossing of the least squares fit then yields the extrapolated value of $\gamma_k^{(e)}$. Note that $\gamma_k^{(e)}$ is close to the ML estimate after just four iterations.

The generalization to the case when p is not known is conceptually easy. In this case, the ML estimates of γ and p are given as the roots of the following equations:

$$\frac{d}{d\gamma} E[\log\, \mathcal{P}_{\gamma,\hat{p}}(X)|Y=y,\hat{\gamma},\hat{p}]\Big|_{\gamma=\hat{\gamma}} = 0$$

$$\frac{d}{dp} E[\log\, \mathcal{P}_{\hat{\gamma},p}(X)|Y=y,\hat{\gamma},\hat{p}]\Big|_{p=\hat{p}} = 0.$$

Similar to the case where p is known, we can now define the vector valued function

$$g(\gamma,p)=\begin{bmatrix} E[u(X,p)/N|Y=y,\gamma,p]-\gamma \\ \dfrac{p^2}{N}\,E\left[\dfrac{d}{dp}\,\dfrac{1}{p}\,u(X,p)\middle|Y=y,\gamma,p\right] \\ +\gamma(1-\log\gamma)+\gamma p^2\,\dfrac{df(p)}{dp} \end{bmatrix}.$$

Then the ML estimates of γ and p are given as the roots of $g(\hat{\gamma},\hat{p})=0$. Note that we can easily compute $g(\gamma_k,p_k)$ when computing the EM updates for γ and p. The computed values of $g(\cdot,\cdot)$ at the past n EM updates are used to obtain least squares fits to two planes. The roots of the fitted planes are then the extrapolated values, $\gamma_k^{(e)}$ and $p_k^{(e)}$.

The nonscalable priors are handled in a similar fashion. In this case, the function $g(\cdot,\cdot)$ is given as

$$g(\sigma,p)=\begin{bmatrix} \dfrac{\sigma}{N}\,E\left[\dfrac{d}{d\sigma}\,u(X/\sigma)|Y=y,p,\sigma\right]+p \\ \dfrac{1}{N}\,E[u(X/\sigma)|Y=y,p,\sigma] \\ -\dfrac{1}{N}\,E[u(X)|p,\sigma=1] \end{bmatrix}.$$

The two components of the vector valued function $g(\cdot,\cdot)$ can be rooted simultaneously or individually to obtain the extrapolated values, $\sigma_k^{(e)}$ and $p_k^{(e)}$.

IV. Experimental Results

In the following two sections, we experimentally study the convergence speed and accuracy of the proposed parameter estimation method. Section IV-A presents results of direct parameter estimation from observed images; while Section IV-B presents results for parameter estimation from incomplete data.

A. Direct Estimation of σ and p

In order to compute the ML estimates of p, we first computed the function $f'(p)$ using (12) and then integrated it using a second-order spline to yield $f(p)$. To compute $f'(p)$, we computed batches of 10 000 full iterations of a 64×64 periodic MRF with an eight-point neighborhood using $b_{i-j}=(2\sqrt{2}+4)^{-1}$ for nearest neighbors and $b_{i-j}=(4\sqrt{2}+4)^{-1}$ for diagonal neighbors. From each batch of 10 000 iterations, an estimate of $f'(p)$ was computed. This procedure was repeated for a single value of p until the estimate was found to stabilize. Then the value of p was updated and the complete procedure repeated. Once the function $f(p)$ was computed, the accuracy of the result was tested by estimating p from sample GGMRF's with known values of p.

Fig. 2 shows the plots of $f(p)$ for $0.4<p<2.0$. Since the computation of $f(p)$ need only be done once, the speed of convergence is not a great issue. However, we found that more iterations were required as p decreased. Also, the function $f'(p)$ was sampled more finely for $p<1$.

Fig. 3 shows a host of natural and synthetic images with their corresponding joint ML estimates of p and σ. Fig. 3(m) and (n) show two sample images generated using the GGMRF model with known values of p and σ. In both cases, the ML estimates are close to the true values.

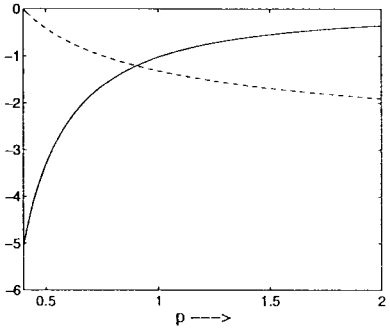


Fig. 2. Solid line shows $f'(p)$ and dashed line shows $f(p)$.

Note that for most natural images except for a few texture images, the ML estimate of p was less than 1.0, and for many images containing man made objects p was less than 0.4. In fact, a similar result has been independently reported by Higdon *et. al.* [35]. Very small values of p may not lead to the best quality MAP reconstructions.[3] In particular, the tomographic cross sections of Fig. 3(o) and (p) yield values of $p<0.4$, which we have found to be undesirable for MAP tomographic reconstruction. Since the ML estimator has well-known optimality properties, this behavior of the ML estimate may be due to the mismatch between the typical tomographic cross sections and the GGMRF model. In light of this result, alternative methods for estimating p, such as those of Jeffs and Pun [39], might be advantageous depending on the intended application.

B. Estimation of σ and p from Incomplete Data

In this section, we study the performance of our proposed algorithms for estimating σ and p from incomplete observations. We present examples using both real and synthetic data for both tomographic reconstruction and image restoration. For the tomographic reconstruction examples, we fix $p=1.1$ and estimate σ because the results of Section IX-A indicate that ML estimates of p from tomographic cross sections are excessively small. However, we emphasize that estimation of σ is a problem of primary importance in tomographic reconstruction, since σ determines the overall smoothness of the reconstruction. MAP image reconstructions are then presented, which indicate that the ML estimate of σ yields a good tradeoff between detail and noise reduction.

To illustrate the utility of our method for optimal joint estimation of p and σ, we apply our method to an image restoration problem, and show that for this case we can accurately estimate p and σ simultaneously from the noisy observations.

For tomographic reconstruction, we found that it was important to restrict the parameter estimates to the support of the object. If the flat background was included, then the estimation of σ tended to be too small and the reconstructions too smooth. For synthetic images, the support was known, but for real

[3] For p less than one, convergence of the MAP estimate can not generally be guaranteed, since the functional being minimized is not convex.

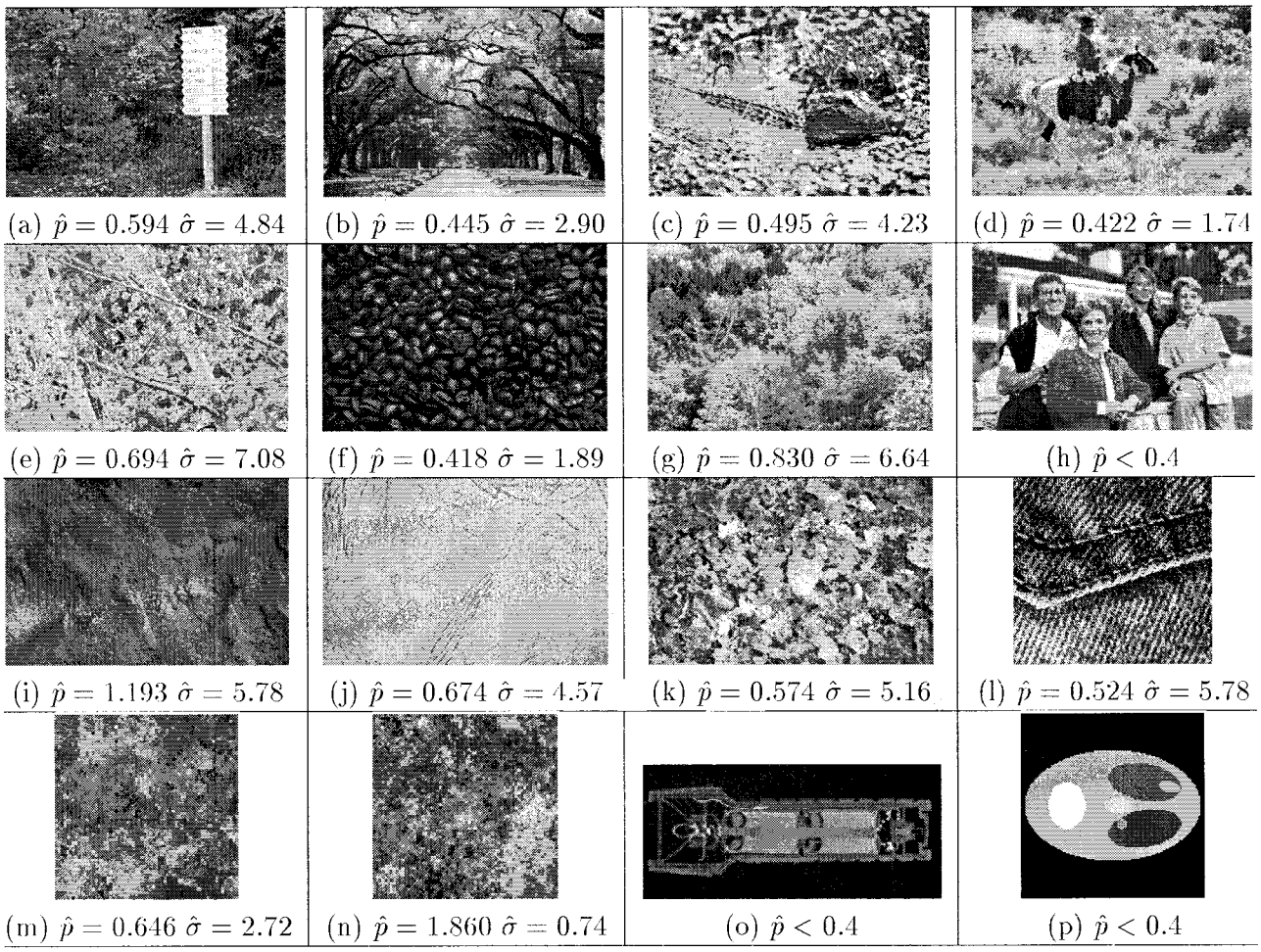
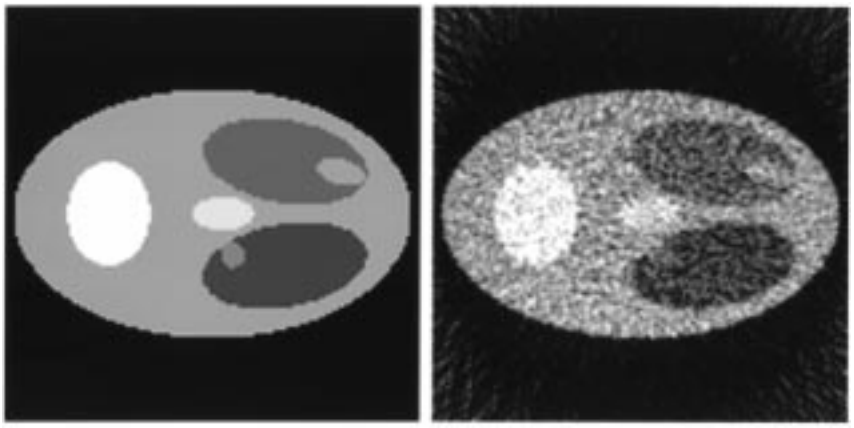


Fig. 3. The above figure shows joint estimation of p and σ for a variety of different images. In particular, (a)–(h) are images of natural scenes, (i)–(l) are images of different textures, (m) and (n) are synthetic images generated from the GGMRF distribution with parameters $p = 0.6, \sigma = 2.34$ and $p = 1.8, \sigma = 0.72$, respectively, (o) is the CBP image obtained from transmission data for a flashlight, and (p) is a synthetic phantom that we will use for emission tomography.



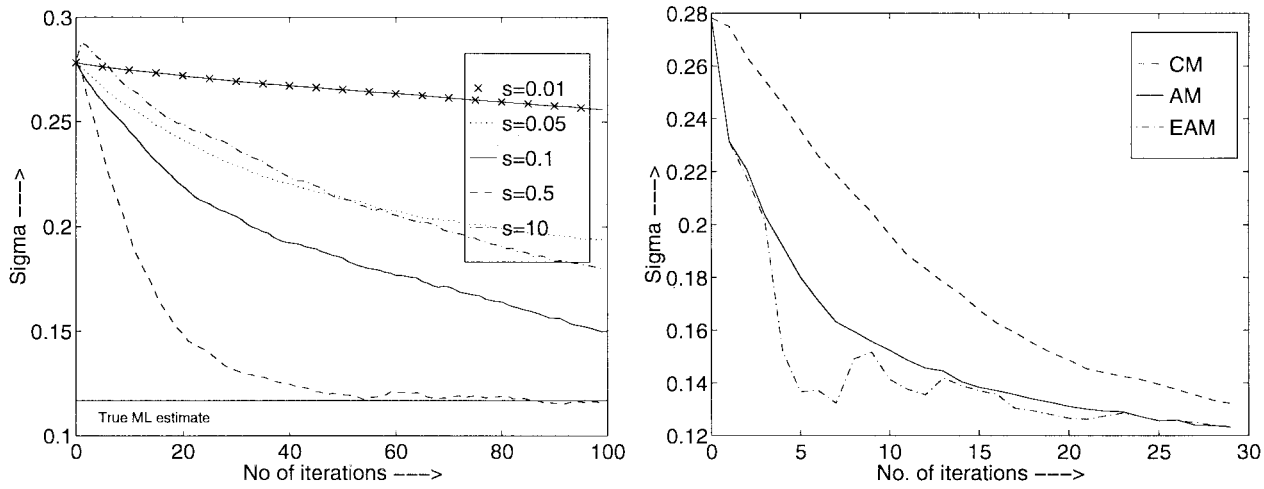
(a) (b)

Fig. 4. (a) Original emission phantom and (b) convolution backprojection (CBP) reconstruction.

images it was extracted by first computing the convolution backprojection (CBP) reconstruction, thresholding with a zero threshold, eroding three times, dilating six times and then eroding three times.

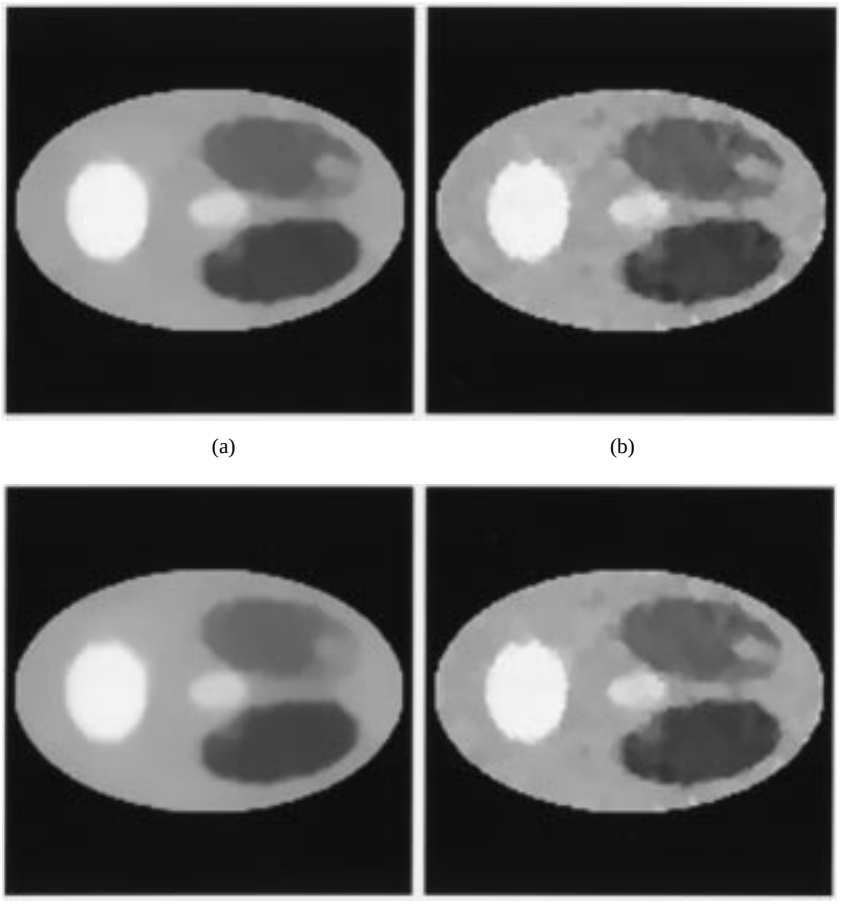
Fig. 4 shows a synthetic emission phantom and the corresponding CBP reconstruction. The emission rates are on an array of 128×128 pixels of size 1.56 mm^2, and 128 Poisson distributed projections are generated at each of 128 uniformly spaced angles. The total photon count was approximately 3 million.

Fig. 5 shows the convergence of the ML estimate for σ using the GGMRF prior. We will refer to the three simulation



(a) (b)

Fig. 5. Convergence plots of σ for the emission phantom modeled by a GGMRF prior $(p = 1.1)$. (a) CM method where s denotes the standard deviation of the symmetric transition distribution. (b) EAM method, AM method, and CM method. All the updates are done using a single sample of X to compute the expectation. However, the true ML estimate is the converged value of σ when 50 samples are used to compute the expectation.

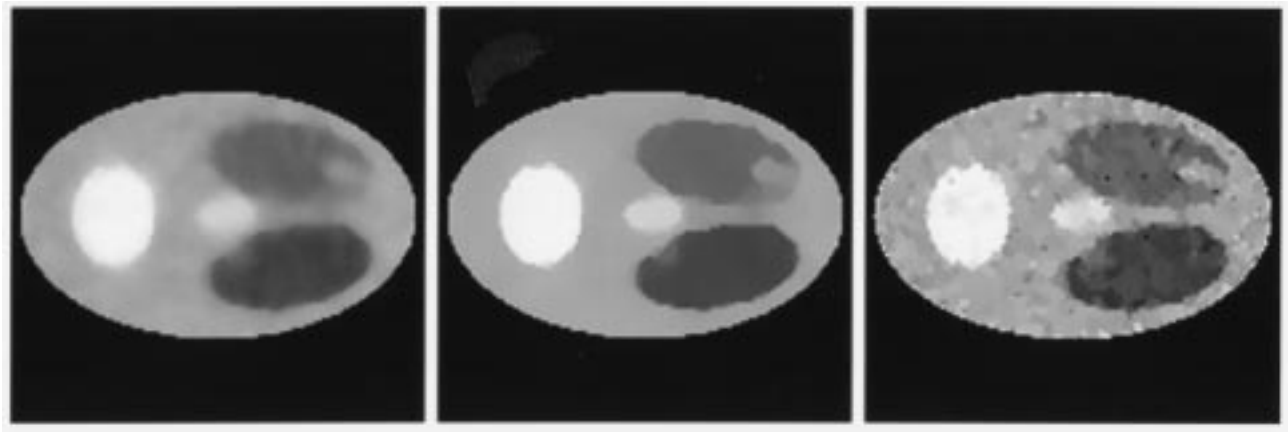


(a) (b)

(c) (d)

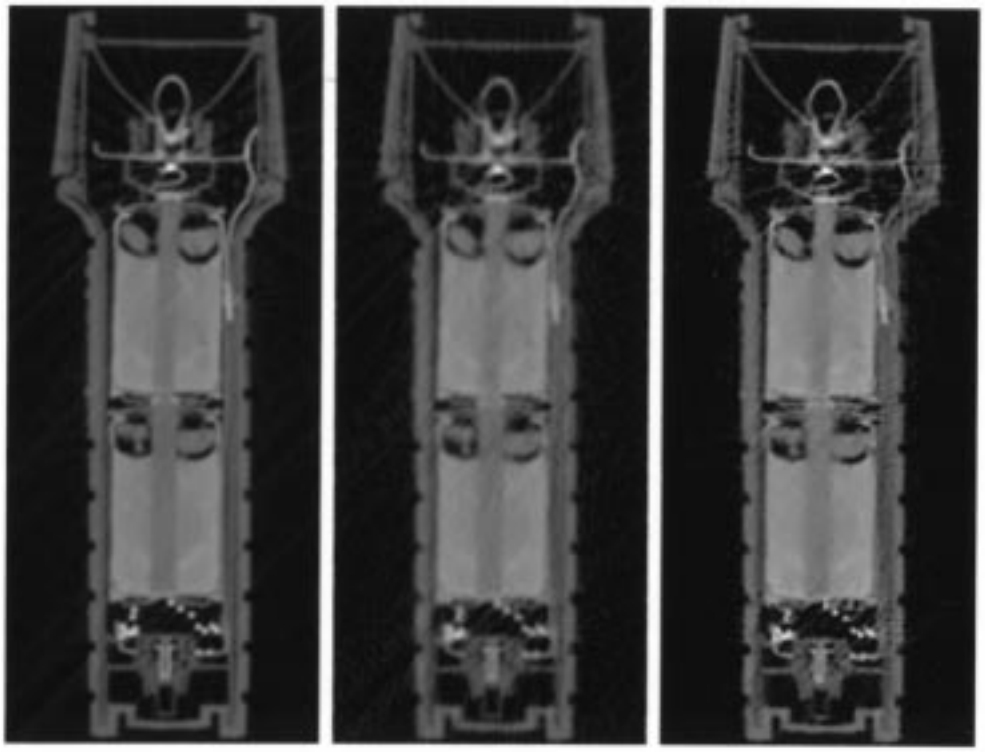
Fig. 6. Reconstructed emission phantom using GGMRF prior with $p = 1.1$ The scale parameter σ is (a) $\hat{\sigma}$, (b) $\hat{\sigma}_{CBP}$, (c) $1/2\hat{\sigma}$, and (d) $2\hat{\sigma}$.

methods as the conventional Metropolis (CM) method, the accelerated Metropolis (AM) method of Section III-C, and the extrapolated-accelerated Metropolis (EAM) method of Section III-D. For each case, each EM update is done after a single full sample of X is computed. The EM algorithm is initialized with the CBP reconstruction and the ML estimate of σ obtained from the CBP reconstruction. Each plot also contains a line labeled as the true ML estimate. This value is computed by using 50 samples of X for each EM update and running the EM updates until convergence.



(a) (b) (c)

Fig. 7. Reconstructed emission phantom using $logcosh(\cdot)$ prior with the scale parameter σ optimally estimated for different values of p. The value of p is (a) 1, (b) 10, and (c) 100.



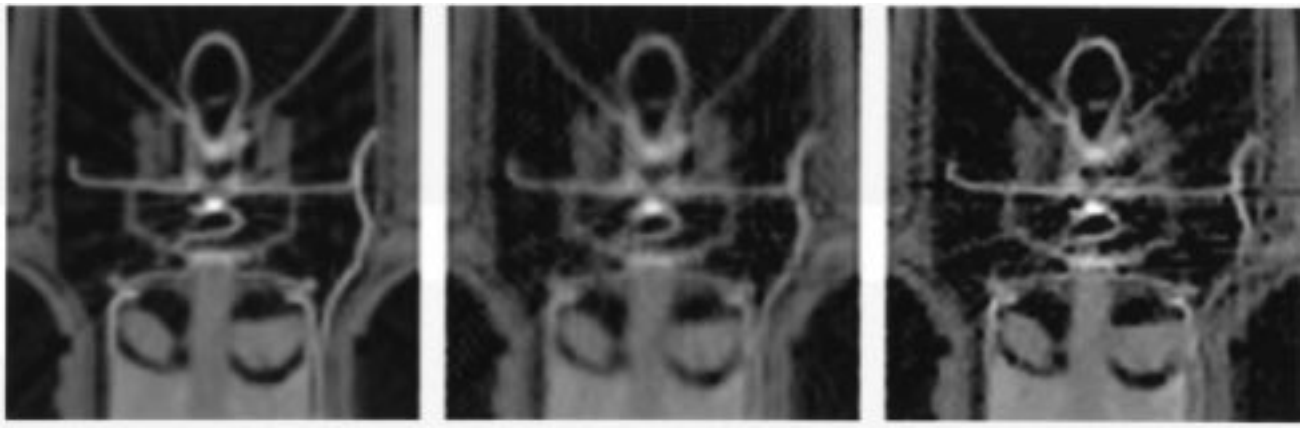
(a) (b) (c)

Fig. 8. (a) Ground truth obtained from high-resolution transmission data, (b) CBP image, and (c) reconstructed image using GGMRF prior with $p = 1.1$ and $\sigma = \hat{\sigma}$. (Data courtesy of T. Neel, Wright-Patterson Air Force Base, and N. Dussausoy, Aracor.)

Fig. 5(a) shows the results using the CM method with a transition distribution chosen to be Gaussian with the variance as the free parameter. Notice that the convergence rate varies substantially with the choice of variance. In practice, it is unclear how to choose the best variance before performing the simulations.

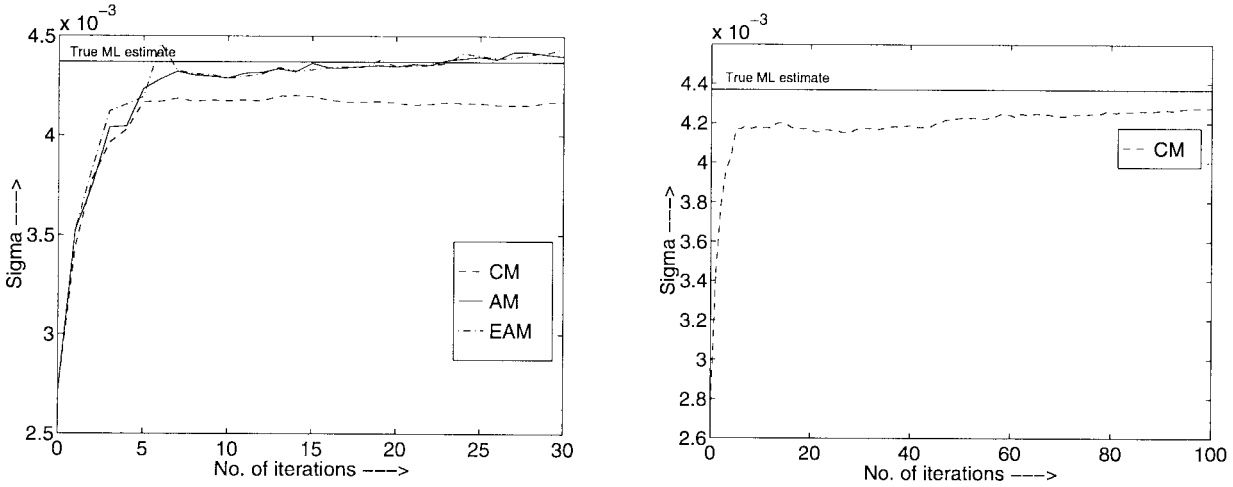
Fig. 5(b) compares the EAM, AM, and CM methods where the CM method uses the variance that produced the most rapid convergence. Notice that the EAM method has the most rapid convergence and all three methods converge to the desired ML value.

Fig. 6 compares the quality of MAP reconstructions using $\hat{\sigma}, \hat{\sigma}/2, 2\hat{\sigma}$, and $\hat{\sigma}_{CBP}$, the estimate obtained directly from the CBP. Of the four results, the ML estimate of σ seems to produce the most desirable tradeoff between detail and noise reduction. Fig. 7 shows the corresponding reconstructions for the $\log\cosh(\cdot)$ prior with $p = 1, 10$, and 100, and the ML estimates of σ. The value $p = 10$ for the $\log\cosh(\cdot)$ prior yields reconstructions similar to that of a GGMRF prior with $p = 1.1$. The value of $p = 100$ for the $\log\cosh(\cdot)$ prior tends to the GGMRF with $p = 1$ and the MAP reconstruction is difficult to compute due to the extremely slow convergence.



(a) (b) (c)

Fig. 9. Blowup of images corresponding to Fig. 8. (a) Ground truth, (b) CBP image, and (c) reconstructed image using GGMRF prior with $p = 1.1$ and $\sigma = \hat{\sigma}$.



(a) (b)

Fig. 10. Comparison of σ's convergence for CM, AM, EAM algorithms for the flashlight phantom modeled by a GGMRF prior $(p = 1.1)$. The true ML estimate is the converged value of σ when 50 samples are used to compute the expectation.

Fig. 8 shows the images corresponding to real transmission data for a flashlight. The original data consisted of 1024 projections taken at 792 equally spaced angles. From this original data, we generated a low resolution data set by retaining every fourth projection at every fourth angle. We used the full resolution data to form a "ground truth" image using CBP reconstruction which is shown in Fig. 8(a). All other reconstructions were then done with the lower resolution data. This approach allows us to determine if the reconstructions using the GGMRF prior actually produce more accurate detail.

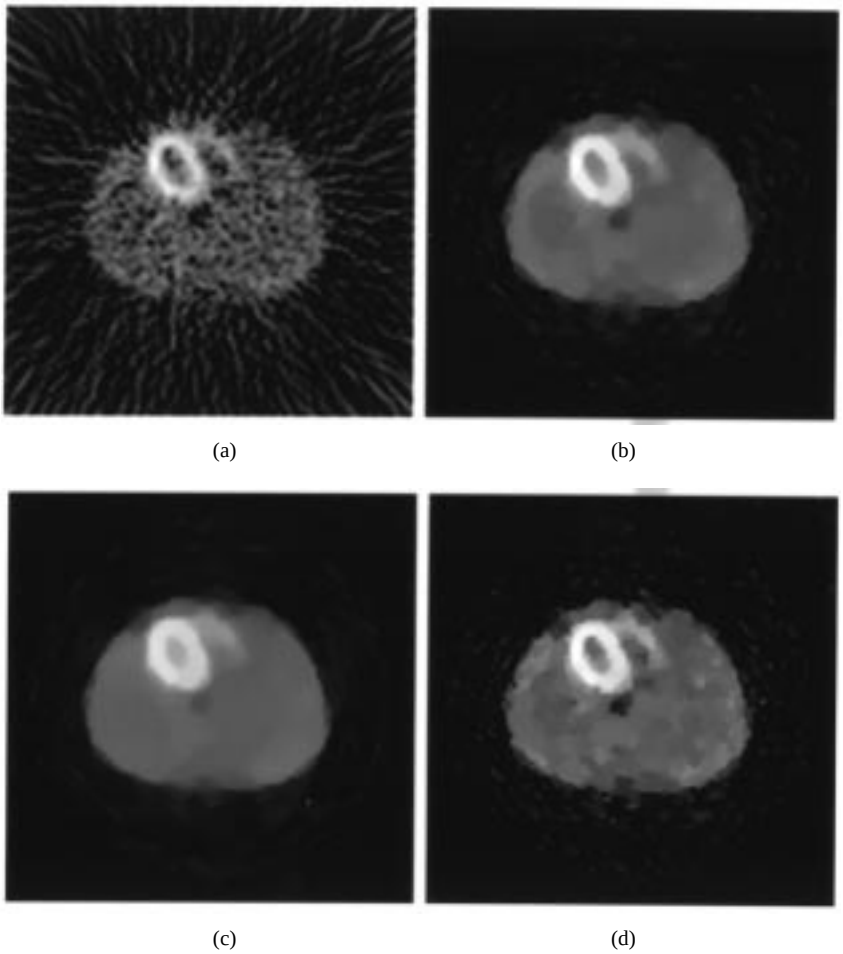
Fig. 8(b) shows the CBP reconstruction and Fig. 8(c) shows the GGMRF reconstruction using the ML estimate of the scale parameter. Fig. 9 shows blowups of the same three images. Notice that the GGMRF reconstruction is sharper than the CBP reconstruction, and in some regions, it contains more detail than the ground truth image reconstructed with 16 times more data.

Fig. 10(a) compares the EM updates of σ for the flashlight data using the CM, AM, and EAM methods. It seems from this plot that the estimate obtained from the CM method has a bias, but Fig. 10(b) shows that after a large number of iterations, the CM method tends toward the ML estimate.

Fig. 11 shows the reconstructions corresponding to a three-dimensional (3-D) SPECT data set obtained from cardiac perfusion imaging using Tc-99m sestamibi. For each slice, 128 projections were taken at 120 uniformly spaced angles between zero and 2π. Fig. 11(a) shows the CBP reconstruction of one slice. The reconstruction was done at 128×128 resolution using 0.356 cm square pixels. The total photon count for this slice was 148 761. Fig. 11(b)–(d) compares the MAP reconstructions corresponding to different values of the scale parameter σ. Again we see that the ML estimate of σ produces a reasonable tradeoff between detail and noise reduction.

Fig. 12 shows the parameter estimation plots using the CM, AM and EAM method for the SPECT data. In this case it takes just one iteration for the AM or EAM method to converge to the ML estimate, whereas the CM method takes about 15 iterations.

Fig. 13(a) shows the original texture image that we use for a restoration example. Fig. 13(b) shows the noisy image obtained by adding uncorrelated Gaussian noise. The SNR of



(a) (b)

(c) (d)

Fig. 11. (a) CBP reconstruction; Reconstructions using GGMRF prior with $p = 1.1$ and (b) $\sigma = \hat{\sigma}$, (c) $\sigma = \hat{\sigma}/2$, (d) $\sigma = 2\hat{\sigma}$. (Data courtesy of T.-S. Pan and M. A. King, University of Massachusetts.)

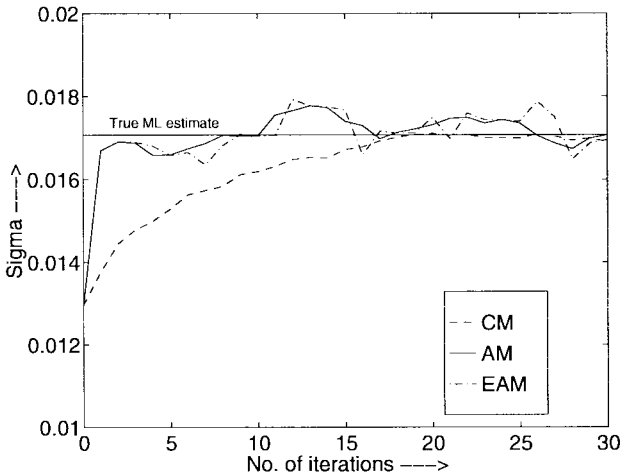


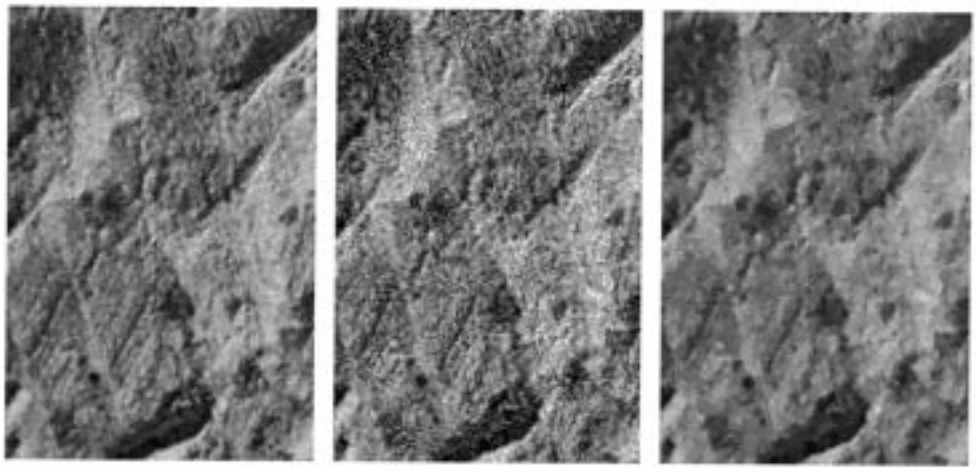
Fig. 12. Comparison of σ's convergence for CM, AM, and EAM algorithms for the SPECT data modeled by a GGMRF prior $(p = 1.1)$. The true ML estimate is the converged value of σ when 50 samples are used to compute the expectation.

the observations is 37 dB. Fig. 14 shows the joint estimation of p and σ for this case using the CM, AM and EAM methods. Note that the ML estimates obtained are very close to the estimates obtained from the original image. It takes about ten iterations for the AM and EAM methods to converge to the ML estimate as compared to 20 iterations for the CM method. Fig. 13(c) shows the MAP restoration for this example using the ML estimates of p and σ.

V. Conclusion

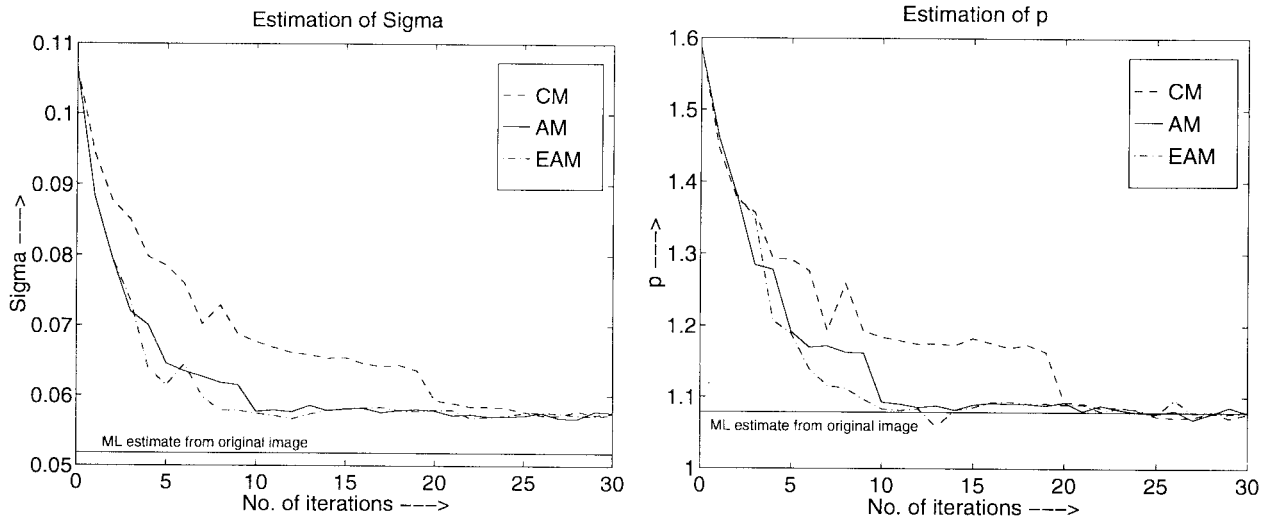
We have shown in this paper that ML estimation of free parameters for Bayesian image reconstruction is feasible for a broad selection of image models and problem settings. Our method is based on parameterization of continuous MRF's by a scale parameter, σ, and a shape parameter, p. For the class of scalable MRF's, the ML estimate of σ may be easily computed in closed form. For other continuous MRF's, the ML estimate of σ may be easily computed as the solution to an equation. Using this result, we also derive a method for computing the ML estimate of the shape parameter, p.

In most practical problems, σ and p must be estimated indirectly from measured data. For this case, we employ the EM algorithm, and develop a fast simulation algorithm together with a method for extrapolating the estimates when the EM algorithm is prematurely terminated. Together these methods allowed good parameter estimates to be computed in fewer than ten iterations for the real and synthetic data sets that were used.



(a) (b) (c)

Fig. 13. (a) Original image. (b) Image corrupted with Gaussian noise (37 dB). (c) Restored image using GGMRF prior and ML estimates of σ and p.



(a) (b)

Fig. 14. These plots show the EM updates for (a) σ, and (b) p for the restoration example using a GGMRF prior. The plots also show the ML estimate obtained for σ and p using the original image. All the updates are done using a single sample of X to compute the expectation.

APPENDIX

In this appendix, we derive the ML estimate of y_T. For the following development, we assume our observations are the random integral projection measurements $\{Z_i = \ln(y_T/Y_i)\}$. Consider the log-likelihood function of z in terms of the unknown dosage parameter y_T. Let the ith actual discretized projection measurement across X be $\tilde{z}_i = A_{i*}x$. Note that Y_i are Poisson distributed with mean and variance $y_T e^{-\tilde{z}_i}$. Then by a simple transformation, we have

$$\mathcal{P}(Z = z|X = x) = \prod_{i=1}^{M} \frac{\exp\{-y_T e^{-\tilde{z}_i}\}(y_T e^{-\tilde{z}_i})^{(y_T e^{-z_i})}}{(y_T e^{-z_i})!} \tag{32}$$

for values of z_i corresponding to positive integer values of y_i and M denotes the number of projections. Stirling's formula provides a simplifying approximation for the factorial, which is relatively accurate for numbers in the typical range of transmission photon counts [57]:

$$(y_T e^{-z_i})! \approx (2\pi y_T e^{-z_i})^{1/2}(y_T e^{-z_i})^{y_T e^{-z_i}} \exp\{-y_T e^{-z_i}\}.$$

Using this substitution, differentiating the logarithm of (32) with respect to y_T and setting the result to zero yields the ML estimate of y_T.

$$\hat{y}_T = \frac{M}{2\sum_{i=1}^{M} [e^{-\tilde{z}_i} - e^{-z_i} + e^{-z_i}(\tilde{z}_i - z_i)]}.$$

ACKNOWLEDGMENT

The authors would like to thank T. Neel, Wright-Patterson Air Force Base, and N. Dussausoy, Aracor, for providing the flashlight data; and T.-S. Pan and M. A. King, University of Massachusetts, for providing the SPECT data.

REFERENCES

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

[2] 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, pp. 194–202, June 1989.

[3] P. J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, pp. 84–93, Mar. 1990.

[4] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, pp. 306–316, Apr. 1984.

[5] J. M. Ollinger, "Maximum-likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm," *IEEE Trans. Med. Imag.*, vol. 13, pp. 89–101, Mar. 1994.

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

[7] B. Hunt, "Bayesian methods in nonlinear digital image restoration," *IEEE Trans. Comput.*, vol. C-26, pp. 219–229, 1977.

[8] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[9] A. Blake and A. Zisserman, *Visual Reconstruction.* Cambridge, MA: MIT Press, 1987.

[10] R. Stevenson and E. Delp, "Fitting curves with discontinuities," in *Proc. 1st Int. Workshop on Robust Computer Vision*, Oct. 1–3, 1990, pp. 127–126.

[11] D. Geman and G. Reynolds, "Constrained parameters and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 367–383, 1992.

[12] C. A. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving map estimation," *IEEE Trans. Image Processing*, vol. 2, pp. 296–310, July 1993.

[13] S. Brette, J. Idier, and A. Mohammed-Djafari, "Scale invariant Markov models for Bayesian inversion of linear inverse problems," in *Maximum Entropy and Bayesian Methods*, J. Skilling and S. Sibisi, Eds. Boston, MA: Kluwer, 1996, pp. 199–212.

[14] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Stat.*, vol. 37, pp. 1554–1563, 1966.

[15] L. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, pp. 164–171, 1970.

[16] L. Shepp and Y. Vardi, "Maximization likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-1, pp. 113–122, Oct. 1982.

[17] D. Pickard, "Asymptotic inference for an ising lattice III: Nonzero field and ferromagnetic states," *J. Appl. Prob.*, vol. 16, pp. 12–24, 1979.

[18] ——, "Inference for discrete Markov fields: The simplest nontrivial case," *J. Amer. Stat. Assoc.*, vol. 82, pp. 90–96, Mar. 1987.

[19] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Stat. Soc. B*, vol. 48, pp. 259–302, 1986.

[20] S. Geman and C. Graffigne, "Markov random field image models and their applications to computer vision," in *Proc. Int. Congr. Mathematicians*, Berkeley, CA, 1986, pp. 1496–1517.

[21] S. Lakshmanan and H. Derin, "Simultaneous parameter estimates and segmentation of Gibbs random fields using simulated annealing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 799–813, Aug. 1989.

[22] B. Chalmond, "An iterative Gibbsian technique for reconstruction of M-ary images," *Pattern Recognit.*, vol. 22, pp. 747–761, 1989.

[23] Y. Ogata, "A Monte Carlo method for an objective Bayesian procedure," *Ann. Inst. Stat. Math.*, vol. 42, pp. 403–433, 1980.

[24] C. Geyer and E. Thompson, "Constrained Monte Carlo maximum likelihood for dependent data," *J. R. Stat. Soc. B*, vol. 54, pp. 657–699, 1992.

[25] B. Gidas, "Parameter estimation for Gibbs distributions from fully observed data," in *Markov Random Fields: Theory and Applications*, R. Chellappa and A. Jain, Eds. New York: Academic, 1993, pp. 471–499.

[26] C. Geyer, "On the convergence of Monte Carlo maximum likelihood calculations," *J. R. Stat. Soc. B*, vol. 56, pp. 261–274, 1994.

[27] G. G. Potamianos and J. K. Goutsias, "Partition function estimation of Gibbs random field images using Monte Carlo simulation," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1322–1332, July 1993.

[28] M. Jerrum and A. Sinclair, "Polynomial-time approximation algorithm for the Ising model," *SIAM J. Comput.*, vol. 22, pp. 1087–1116, 1993.

[29] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. Image Processing*, vol. 1, pp. 322–336, July 1992.

[30] A. Mohammad-Djafari, "On the estimation of hyperparameters in Bayesian approach of solving inverse problems," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Minneapolis, MN, Apr. 27–30, 1993, pp. 495–498.

[31] K. Lange, "An overview of Bayesian methods in image reconstruction," *Proc. SPIE Conf. Digital Image Synthesis and Inverse Optics*, San Diego, CA, 1990, vol. SPIE-1351, pp. 270–287.

[32] Z. Zhou, R. Leahy, and J. Qi, "Approximate maximum likelihood hyperparameter estimation for Gibbs priors," *IEEE Trans. Image Processing*, vol. 6, pp. 844–861, June 1997.

[33] A. Mohammad-Djafari, "Joint estimation of parameters and hyperparameters in a Bayesian approach of solving inverse problems," in *Proc. IEEE Int. Conf. Image Processing*, Lausanne, Switzerland, Sept. 16–19, 1996, vol. II, pp. 473–476.

[34] R. Schultz, R. Stevenson, and A. Lumsdaine, "Maximum likelihood parameter estimation for non-Gaussian prior signal models," *Proc. IEEE Int. Conf. Image Processing*, Austin, TX, Nov. 1994, vol. 2, pp. 700–704.

[35] D. M. Higdon, *et al.*, "Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data," *IEEE Trans. Med. Imag.*, vol. 16, Oct. 1997.

[36] Z. Zhou and R. Leahy, "Approximate maximum likelihood hyperparameter estimation for Gibbs priors," in *Proc. IEEE Int. Conf. Image Processing*, Washington, DC, Oct. 23–26, 1995, pp. 284–287.

[37] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Trans. Signal Processing*, vol. 40, pp. 2570–2583, Oct. 1992.

[38] W. Pun and B. Jeffs, "Shape parameter estimation for generalized Gaussian Markov random field models used in MAP image restoration," in *Proc. 29th Asilomar Conf. Signals, Systems, and Computers*, Oct. 29–Nov. 1, 1995.

[39] B. D. Jeffs and W. H. Pun, "Simple shape parameter estimation from blurred observations for a generalized Gaussian MRF image prior used in MAP image restoration," in *Proc. IEEE Int. Conf. Image Processing*, Lausanne, Switzerland, Sept. 16–19, 1996, pp. 465–468.

[40] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Int. Stat. Inst.*, vol. LII-4, pp. 5–21, 1987.

[41] C. A. Bouman and K. Sauer, "Maximum likelihood scale estimation for a class of Markov random fields," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Adelaide, Australia, Apr. 19–22, 1994, vol. 5, pp. 537–540.

[42] C. A. Bouman, K. Sauer, and S. S. Saquib, "Tractable models and efficient algorithms for Bayesian tomography," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, May 9–12, 1995, Detroit, MI, pp. 2907–2910.

[43] S. S. Saquib, C. A. Bouman, and K. Sauer, "Efficient ML estimation of the shape parameter for generalized Gaussian MRF," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Atlanta, GA, May 7–10, 1996, vol. 4, pp. 2229–2232.

[44] ——, "ML parameter estimation for Markov random fields, with applications to Bayesian tomography," Tech. Rep. TR-ECE 95-24, School Electr. Comput. Eng., Purdue Univ., West Lafayette, IN, Oct. 1995.

[45] K. Sauer and C. A. Bouman, "Maximum likelihood dosage estimation for Bayesian tomography," in *Proc. IEEE Int. Conf. on Image Processing*, Austin, TX, Nov. 13–16, 1994, vol. 2, pp. 844–848.

[46] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, pp. 1087–1091, 1953.

[47] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[48] P. H. Peskun, "Optimum Monte-Carlo sampling using Markov chains," *Biometrika*, vol. 60, pp. 607–612, 1973.

[49] P. J. Green and X. Liang Han, "Metropolis methods, Gaussian proposals and antithetic variables," in *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis*, P. Barone, A. Frigessi, and M. Picconi, Eds. Berlin, Germany: Springer-Verlag, 1992, pp. 142–164.

[50] C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Image Processing*, vol. 5, pp. 480–492, Mar. 1996.

[51] A. Blake, "Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 2–30, Jan. 1989.

[52] J. Besag, "Toward Bayesian image analysis," *J. Appl. Stat.*, vol. 16, pp. 395–407, 1989.

[53] J. A. O'Sullivan, "Divergence penalty for image regularization," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Adelaide, Australia, Apr. 19–22, 1994, vol. 5, pp. 541–544.
[54] S. Ihara, *Information Theory for Continuous Systems.* Singapore: World Scientific, 1993.
[55] Y. Ogata and M. Tanemura, "Likelihood analysis of spatial point patterns," *J. R. Stat. Soc. B*, vol. 46, pp. 496–518, 1984.
[56] ———, "Likelihood estimation of soft-core interaction potentials for Gibbs point patterns," *Ann. Inst. Stat. Math.*, vol. 41, pp. 583–600, 1989.
[57] M. Abramowitz and I. Stegun, eds., *Handbook of Mathematical Functions.* New York: Dover, 1965.

Suhail S. Saquib was born in Raipur, India, in 1970. He received the B.S degree in electronics and electrical communication engineering from Indian Institute of Technology, Kharagpur, India, in 1991, and the M.S and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1992 and 1997, respectively.

During the summer of 1996, he was an intern at Los Alamos National Laboratories, where he worked in the area of medical optical tomography. He is currently working in the Image Science Laboratory, Polaroid Corporation, Cambridge, MA. His interests include computerized tomography, computer vision, statistical modeling, and pattern recognition.

Charles A. Bouman (S'86–M'89–SM'97) 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, respectively, he received the M. A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, under the support of an IBM graduate fellowship.

From 1982 to 1985, he was a Staff Member in the Analog Device Technology Group, Massachusetts Institute of Technology Lincoln Laboratory, Lexington. In 1989, he joined the faculty of Purdue University, West Lafayette, IN, where he currently holds the position of Associate Professor in the School of Electrical and Computer Engineering. His 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 browsing, tomographic reconstruction, document segmentation, and image rendering. He has performed research for numerous government and industrial organization including National Science Foundation, the U.S. Army, Hewlett-Packard, NEC Corporation, Apple Computer, Xerox, and Eastman Kodak.

Dr. Bouman is a Member of SPIE and IS&Y professional societies. From 1991 through 1993, he was also an NEC Faculty Fellow. 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 ON IMAGE PROCESSING. He is currently a member of the IEEE Image and Multidimensional Signal Processing Technical Committee.

Ken Sauer (S'85–M'89) was born in Decatur, IN. He received the B.S.E.E. degree in 1984 and the M.S.E.E. degree in 1985 from Purdue University, West Lafayette, IN, 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. He is involved in research of statistical methods for tomographic image estimation and other nondestructive evaluation problems, and stochastic image modeling.