

10-1-1995

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

Suhail S. Saquib

Purdue University School of Electrical and Computer Engineering

Charles A. Bouman

Purdue University School of Electrical and Computer Engineering

Ken Sauer

University of Notre Dame Department of Electrical Engineering

Follow this and additional works at: <http://docs.lib.purdue.edu/ecetr>

Saquib, Suhail S.; Bouman, Charles A.; and Sauer, Ken, "ML Parameter Estimation for Markov Random Fields, with Applications to Bayesian Tomography" (1995). *ECE Technical Reports*. Paper 145.

<http://docs.lib.purdue.edu/ecetr/145>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

ML PARAMETER ESTIMATION FOR
MARKOV RANDOM FIELDS, WITH
APPLICATIONS TO BAYESIAN
TOMOGRAPHY

SUHAIL S. SAQUIB
CHARLES A. BOUMAN
KEN SAUER

TR-ECE 95-24
OCTOBER 1995



SCHOOL OF ELECTRICAL
AND COMPUTER ENGINEERING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907-1285

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

Suhail S. Saquib, Charles A. Bouman
School of Electrical Engineering
Purdue University
West Lafayette, IN 47907.
{saquib, bouman}@ecn.purdue.edu

Ken Sauer
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556.
sauer.1@nd.edu

*This work was supported by National Science Foundation grant number MIP93-00560.

Abstract

Markov random fields (MRF) have proven useful for modeling the *a priori* information in Bayesian tomographic reconstruction problems. However, optimal parameter estimation of the MRF model remains a difficult problem due to the intractable nature of the partition function. In this report, we propose a fast parameter estimation scheme to obtain optimal estimates of the free parameters associated with a general MRF model formulation. In particular, for the generalized Gaussian MRF (GGMRF) case, we show that the ML estimate of the temperature T has a simple closed form solution. We present an efficient scheme for the ML estimate of the shape parameter p by an off-line numerical computation of the log of the partition function. We show that this approach can be extended to compute the parameters associated with a general MRF model. In the context of tomographic reconstruction, the difficulty of the ML estimation problem is compounded by the fact that the parameters depend on the unknown image. The EM algorithm is used to solve this problem. We derive fast simulation techniques for efficient computation of the expectation step. We also propose a method to extrapolate the estimates when the simulations are terminated prematurely prior to convergence. Experimental results for the emission and transmission case show that the proposed methods result in substantial savings in computation and superior quality images.

1 Introduction

In the past decade, Bayesian methods have become popular in tomographic image reconstruction [1] and restoration problems [2]. The objective of these methods is to use both a model for the observations and a model for the unknown image in the estimation process. Markov random fields (MRF) have proven useful for modeling the *a priori* information in the tomographic setting. MRFs have been applied extensively in emission tomography [3, 4, 5] and to a lesser extent in transmission tomography [6, 7]. The MRF model is equivalent to a Gibbs distribution and is often specified in terms of a potential function which assigns a cost to differences between neighboring pixels. The preponderance of the previous work has focused primarily on the quadratic choice for the potential function or Gaussian MRF [8]. Although this particular choice has many analytical advantages, reconstructed edges may be blurred due to the excessive cost assigned to abrupt transitions. Many alternative potential functions have been proposed in the literature which help to alleviate this problem [9, 4, 5, 10, 11, 12, 13]. Among these are the non-convex function of Blake and Zisserman [9] and Geman and Reynolds [12], the logcosh function of Green [5], the Huber function of Stevenson and Delp [10], and the generalized Gaussian function of Bouman and Sauer [13]. In particular, the generalized Gaussian MRF (GGMRF) uses a potential function similar to the log of the generalized Gaussian noise density found commonly in robust detection and estimation [14]. It renders edges accurately without prior knowledge of their size, and it often results in a convex optimization problem with a unique global minimum [13].

The stochastic models for the observed data and the image have certain free parameters associated with them. The parameters of the data model often describe the amount of measurement noise and the parameters of the image model often describe edge behavior and image variation. For completely unsupervised reconstruction and restoration, these free parameters of the image and data models have to be estimated from the data itself. Unsupervised estimation of these parameters will be the focus of this report.

In the past, many authors have grappled with the problem of estimating the smoothing parameter in regularized image restoration using a quadratic regularization criteria. This is similar to using the Gaussian MRF to model the image in the Bayesian framework. Consequently, a variety of techniques have to been proposed to solve this problem such as constrained least squares [15], equivalent degrees of freedom [16], predicted mean square error [17], cross validation [18], and the expectation-maximization (EM) algorithm for ML estimation [19]. For a review of these methods see [20, 21].

We are more interested in the non-Gaussian (i.e, not quadratic) cases since they can better

characterize natural images¹. However, parameter estimation for a non-Gaussian MRF model remains a difficult problem, due to the intractable nature of the partition function. Owing to this difficulty, a host of approximation methods have been suggested in the literature [22, 23, 24]. In particular, the coding [22] and pseudolikelihood [23] method of Besag has been applied extensively in the discrete case [25, 26, 27].

The difficulty of parameter estimation in the reconstruction or restoration problems is compounded by the parameters' dependence on the unknown image. The EM algorithm [28] is an elegant method to address problems of this nature. EM hypothesizes the existence of a set of complete data from which ML estimation would be feasible, then attempts to iteratively maximize over the expectation of the complete data. As the name suggests, it consists of two steps - expectation and maximization. The maximization step is generally trivial if the ML parameter estimates are known as functions of the complete data. On the other hand, the expectation step may be intractable, requiring stochastic simulation for its evaluation.

Using the EM framework, Zhang [29] proposed an ML estimation scheme where mean field theory is used to approximate the involved expectation. More recently, Schultz et al. [30] suggested an ML estimation scheme where they approximate the posterior distribution with a quadratic function to compute the log-likelihood of the observations explicitly as a function of the parameters. The ML estimates are obtained by maximizing the log-likelihood with respect to the parameters. Pun and Jeffs [31] suggested a data directed procedure to estimate the shape parameter of a GGMRF by computing the kurtosis of differences between neighboring pixels.

Surprisingly, the partition function for a GGMRF is tractable with respect to the temperature T . This result was derived recently by Bouman and Sauer [32] when they showed that the exact ML estimate of T has a simple closed form solution. The same result was derived earlier by Lange [33], who observed that joint estimation of the image and the temperature T results in a divergent estimator for the transmission case.

Geman and McClure [3] suggested an EM procedure to compute the exact ML estimate of the temperature T associated with a general MRF. The intractable expectation involved in the EM algorithm was computed by using the Metropolis simulation method, and the maximization step was made trivial by an off-line calculation of the expected energy with respect to the prior distribution of the image. They estimated only the temperature T in this fashion, leaving the scaling parameter to be hand picked. To circumvent the computational demands of the EM algorithm, they also suggested a direct estimation method for T based

¹In fact we show later on, by estimating the shape parameter of a GGMRF, that most natural images appear more Laplacian than Gaussian.

on the observations without requiring the intermediate reconstruction of the unknown image. They called this approximate procedure the method of moments [3]. Ogata [34] suggested a ML estimation scheme that can be applied to essentially any MRF. In essence, his method works by numerically computing the log of the partition function. Though this method is very general, it is also very computationally intensive.

In this report, we present efficient algorithms for computing maximum likelihood parameter estimates for MRF models used in Bayesian reconstruction and restoration problems. For the GGMRF case, we show that the closed form ML estimate of \mathbf{T} and an off-line numerical computation of the log of the partition function allow the estimate of the shape parameter to be obtained through a simple one dimensional optimization. We show that this method can be extended to optimally estimate the parameters of any general MRF model. We also compute the ML estimate of input dosage for the transmission case where such information might have been lost during the collection process. Since the parameters depend on the unknown image, we use the EM framework to compute the estimates, with stochastic simulation replacing the intractable expectation step.

We propose a fast simulation technique based on the Metropolis algorithm [35] for efficient computation of the expectation step. We also introduce a method to extrapolate the parameter estimates when the simulations are terminated prematurely prior to convergence. Experimental results for the emission and transmission case show that the proposed methods result in substantial savings in computation and superior quality images.

The organization of this report is as follows: Section 2 outlines the stochastic data models used in emission and transmission tomography; Section 3 discusses the different prior models and derives the ML estimate of the parameters for these models; Section 4 outlines the EM framework; Section 5 and 6 derives the fast simulation technique and the extrapolation method respectively; Section 7 presents the experimental results while section 8 provides the concluding remarks.

2 Stochastic Data Models for Tomography

In this section we introduce the notation and the stochastic data models used in emission and transmission tomography. We will use upper case letters to denote random vectors and lower case letters to denote a particular realization of the random vector.

Let \mathbf{x} denote the set of emission intensities in the emission case and the set of attenuation densities in the transmission case. For the emission case, let A_{ij} be the probability that a photon emitted from cell \mathbf{j} is registered at the i^{th} detector. Let \mathbf{A} be the matrix with elements $\{A_{ij}\}$, and let \mathbf{A}_i denote the vector formed by its i^{th} row. Let \mathbf{y} denote the set

of measurements of Poisson-distributed photon counts at the detectors for all angles and displacements. Then the conditional distribution of the photon counts Y given x for the emission case is

$$\mathcal{P}(Y = y|x) = \prod_i \frac{\exp\{-A_{i^*}x\} \{A_{i^*}x\}^{y_i}}{y_i!} \quad (1)$$

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. Using (1), the log-likelihood may be computed as

$$\text{(emission) } L(Y = y|x) = \sum_i (-A_{i^*}x + y_i \log\{A_{i^*}x\} - \log(y_i!)) \quad (2)$$

The transmission case is similar, but has A_{ij} corresponding to the length of intersection between the j^{th} cell and the i^{th} projection. Let the input photon counts be Poisson-distributed with rate y_T . The conditional log-likelihood of Y given x for the transmission case is

$$\text{(transmission) } L(Y = y|x) = \sum_i (-y_T e^{-A_{i^*}x} + y_i(\log y_T - A_{i^*}x) - \log(y_i!)) \quad (3)$$

In the Bayesian estimation problem, the unknown image X is assumed to be a random field with some prior distribution. Let $\mathcal{P}(x)$ denote the prior distribution of X . Then the complete model for the transmission and emission case is given as

$$\log \mathcal{P}(y, x) = L(Y = y|x) + \log \mathcal{P}(x)$$

The maximum a posteriori (MAP) estimate is commonly used to reconstruct the unknown image in tomographic reconstruction and restoration problems.

$$\hat{X}_{MAP} = \arg \max_{x > 0} \{L(Y = y|x) + \log \mathcal{P}(x)\}$$

In this report, we will be using the MAP estimate for all our reconstructions. The iterative coordinate descent (ICD) method [6] will be used to compute the MAP estimate. For a more detailed derivation of these models see for example [36].

3 Prior Models

We will be modeling X as an MRF or, equivalently, a Gibbs distribution with the following form.

$$\mathcal{P}(X = x) = \begin{cases} \frac{1}{Z} \exp\left\{-\frac{1}{T}u(x, \theta)\right\} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where we have imposed the positivity constraint on X ; a reasonable assumption for emission and transmission tomography. $z(\cdot, \cdot)$ is the normalizing constant of the distribution known as the partition function. T is the temperature and β is a general parameter whose role depends on the choice of the energy function $u(\cdot, \cdot)$. Let $\beta = \mathbf{a}$ when it takes on the role of a scaling parameter, and $\beta = \mathbf{p}$ when it takes on the role of a shape parameter. We will be considering $u(\cdot, \cdot)$ of the form

$$u(x, \theta) = \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \rho(x_i - x_j, \theta) \quad (5)$$

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.

Many previous studies have chosen $\rho(\Delta, \mathbf{a}) = |\Delta/\sigma|^2$ [8]. Although this particular choice has many analytical advantages, it generally results in reconstructions which are either excessively noisy or blurred. On the other hand, non-quadratic functions are interesting because they can potentially model both the edges and smooth regions of images. A typical non-quadratic function which is also non-convex was proposed by Blake and Zisserman [9]

$$\rho(\Delta, \sigma) = \min \left\{ \left| \frac{\Delta}{\sigma} \right|, 1 \right\}^2$$

This function is quadratic for $A < \mathbf{a}$, but the flat region for $A > \mathbf{a}$ allows sharp edges to form in the reconstructed image.

For the purposes of modeling images, this distribution has some significant practical and theoretical disadvantages. Since the function is nonconvex it is generally impractical to globally minimize. Consequently, the MAP estimate can only be approximated and the solution achieved usually depends substantially on the method used to perform the minimization. Also, the reconstructed image using this prior has an unnatural quality since edges with magnitude greater than \mathbf{a} are sharp while those of magnitudes lower than \mathbf{a} are smooth. In addition, it was shown in [13] that the reconstructed image is not continuously dependent on the observations. Therefore a small change in the observation can result in a substantially different reconstruction.

We can overcome the above disadvantages by choosing $\rho(\cdot, \cdot)$ as a convex function. A convex choice for $\rho(\cdot, \cdot)$ makes the overall cost function for the MAP estimate convex and consequently the reconstruction problem becomes computationally feasible. Several non-quadratic convex functions for $\rho(\cdot, \cdot)$ have been suggested in the literature [5, 10, 11, 13]. For example, Green [5] employed a function of the form

$$\rho(A, \mathbf{a}) = \log \cosh \left(\frac{\Delta}{\sigma} \right) \quad (6)$$

This potential function is approximately quadratic when A is less than a and approximately linear when A is greater than a . Stevenson and Delp [10] chose the Huber function first introduced in robust statistics [37]. This function is similar to Green's function since it is quadratic below a certain threshold and linear above it. Bouman and Sauer [13] suggested the generalized Gaussian prior

$$\rho(\Delta, p) = |\Delta|^p \quad (7)$$

where p lies between 1 and 2. This particular choice of $\rho(\cdot, \cdot)$ makes the energy function $u(\cdot, \cdot)$ scalable i.e for all $x \in \mathbb{R}^N$ and $\alpha > 0$

$$u(\alpha x, p) = \alpha^p u(x, p) \quad (8)$$

where p is some positive constant. The consequence of the scaling property of the energy function is that the temperature parameter T becomes equivalent to the scale parameter. In fact, an alternate parameterization obtained by substituting $T = pa^p$ yields more intuitively appealing results since the scale parameter a is physically meaningful. However, T is more convenient to use in the derivations since it simplifies the computation. Therefore, we will perform all the intermediate steps in the derivations using T and present the final result in terms of a by substituting $T = pa^p$.

Fig. 1 shows how the different priors compare to each other. Notice that the $\text{logcosh}(\cdot)$ and the Huber function are similar in form. We will therefore restrict ourselves to the $\text{logcosh}(\cdot)$ prior and the generalized Gaussian prior in our subsequent treatment since we believe these are representative. However, the method we present will be adaptable to other models.

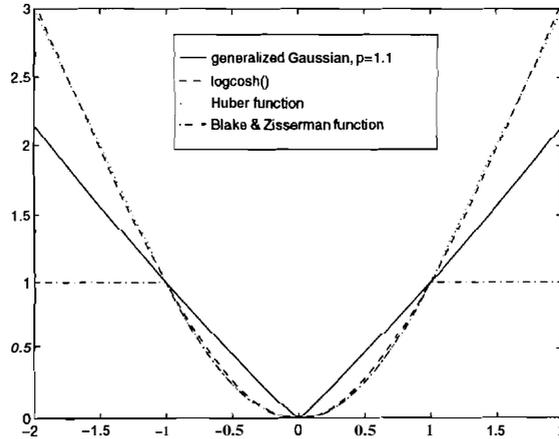


Figure 1: The above plot shows how the different priors compare to each other.

3.1 ML Estimate of a for Generalized Gaussian Prior

In this section, we derive the exact closed form ML estimate of a for a GGMRF. The derivation proceeds on similar lines as Lange [33], but we also impose the positivity constraint on X .

As noted above, the energy function of a GGMRF is scalable. The following theorem shows that the partition function $z(T, p)$ can be explicitly expressed as a function of T .

Theorem *For the class of MRF with scalable energy functions and $x \in \mathbb{R}^N$ or $x \in \mathbb{R}^{+N}$, the partition function can be expressed as $z(p\sigma^p, p) = \sigma^N p^{N/p} z(1, p)$.*

Proof:

$$\begin{aligned}
 z(p\sigma^p, p) &= \int_{x \in \mathbb{R}^{+N}} \exp \left\{ -\frac{1}{p\sigma^p} u(x, p) \right\} dx \\
 &= \int_{x \in \mathbb{R}^{+N}} \exp \left\{ -u(x/(\sigma p^{1/p}), p) \right\} dx \\
 &= \sigma^N p^{N/p} \int_{x \in \mathbb{R}^{+N}} \exp \left\{ -u(x, p) \right\} dx \\
 &= \sigma^N p^{N/p} z(1, p)
 \end{aligned} \tag{9}$$

The same proof applies when $x \in \mathbb{R}^N$.

Using (9), we can rewrite the normalized log-likelihood of x as

$$\frac{-1}{N} \log \mathcal{P}(x) = \frac{1}{Np\sigma^p} u(x, p) + \log \sigma + \frac{\log p}{p} + \frac{1}{N} \log z(1, p) \tag{10}$$

Differentiating (10) with respect to a and setting the result to zero yields the ML estimate of σ as

$$\hat{\sigma}(x, p) = \sqrt[p]{\frac{1}{N} u(x, p)} \tag{11}$$

We note that for the i.i.d. Gaussian case $u(x) = \sum_i x_i^2$, and this reduces to the familiar expression

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_i x_i^2}$$

3.2 Joint ML Estimate of a and p for Generalized Gaussian Prior

In this section, we will derive a method for computing the joint ML estimate of p and σ for the GGMRF model. Using (10), we express the joint ML estimate as

$$(\hat{\sigma}, \hat{p}) = \arg \min_{(\sigma, p)} \left\{ \frac{1}{Np\sigma^p} u(x, p) + \log \sigma + \frac{\log p}{p} + \frac{1}{N} \log z(1, p) \right\} \tag{12}$$

We can reduce this problem to a one dimensional optimization since we have the closed form ML estimate of σ (11) in terms of p . Substituting (11) in (12) we obtain the joint ML estimate as

$$\hat{p}(x) = \underset{a}{\min} \left\{ \log \hat{\sigma}(x, p) + \frac{1}{P} + \frac{\log p}{P} + f(p) \right\} \quad (13)$$

$$\hat{\sigma}(x) = \hat{\sigma}(x, \hat{p}(x)) \quad (14)$$

where

$$f(p) = \frac{1}{N} \log z(1, p) = \frac{1}{N} \log \int_{x \in \mathbf{R}^{+N}} \exp \{-u(x, p)\} dx$$

The intuitive meaning of this result becomes clearer if we examine the form of (13). We notice that the variation of scale σ with respect to p is a sufficient statistic to determine the ML estimate of p .

Direct computation of $f(p)$ would require the numerical evaluation of an N dimensional integral which is not feasible. Instead a more elegant method of computing $f(p)$ is through its derivative.

$$\begin{aligned} f'(p) = \frac{d}{dp} f(p) &= \frac{-1}{Nz(1, p)} \mathbf{J}_{x \in \mathbf{R}^{+N}} \frac{d}{dp} u(x, p) \exp \{-u(x, p)\} dx \\ &= \frac{-p}{N} \sum_{\{i, j\} \in \mathcal{N}} b_{i-j} E \left[|X_i - X_j|^{p-1} \mid \sigma^p = 1/p, p \right] \end{aligned} \quad (15)$$

We will show in section 5 that a fast simulation method can be used to numerically compute this expectation. The function $f'(p)$ is computed off-line prior to the estimation procedure. The normalization by N is essential for the function to be useful for any image size. A second order spline is used to interpolate the $f'(p)$. The fitted spline is then integrated to obtain $f(p)$. Table 1 shows the computed values of $f'(p)$ and $f(p)$ for an 8 point neighborhood and Figure 2 shows plots of $f'(p)$ and $f(p)$.

The ML estimate of p is obtained by computing (13) for a finely spaced set of values for $0.8 \leq p \leq 2$ and finding the minimum with respect to p . (We could also reduce this computation by employing a fast rooting method such as the half interval search to root the derivative of $\log \mathcal{P}(x)$ with respect to p .) Figure 3(a) shows a transmission phantom and figure 3(b) shows a natural image with their corresponding negative log-likelihood as a function of p . For both these cases, the estimate hits the constraint of 0.8. In fact, this was the case for most of the natural images that we tried. Figure 3(c) is the Gaussian noise corrupted version of Figure 3(b). We observe that this raises the ML estimate of p to 1.4.

p	$f'(p)$	$f(p)$	p	$f'(p)$	$f(p)$
0.80	-1.7990	0	1.45	-0.3985	-0.5506
0.85	-1.5381	-0.0834	1.50	-0.3668	-0.5698
0.90	-1.3240	-0.1550	1.55	-0.3313	-0.5872
0.95	-1.1402	-0.2166	1.60	-0.3076	-0.6032
1.00	-0.9968	-0.2700	1.65	-0.2901	-0.6181
1.05	-0.8788	-0.3169	1.70	-0.2643	-0.6320
1.10	-0.7766	-0.3583	1.75	-0.2483	-0.6448
1.15	-0.6916	-0.3950	1.80	-0.2318	-0.6568
1.20	-0.6170	-0.4277	1.85	-0.2169	-0.6680
1.25	-0.5554	-0.4570	1.90	-0.2037	-0.6785
1.30	-0.5061	-0.4835	1.95	-0.1915	-0.6884
1.35	-0.4645	-0.5078	2.00	-0.1819	-0.6978
1.40	-0.4249	-0.5301			

Table 1: The computed value for $f(p)$ and $f'(p)$ are listed. Second order neighborhood with periodic boundary conditions was used for the computation. The weights b_{i-j} for each pixel i was set to 0.1464 and 0.1036 for nearest and diagonal neighbors respectively.

3.3 ML estimate of T and σ for $\text{logcosh}(\cdot)$ Prior

In this section, we show that the optimal estimate for T and a can be obtained for non-scalable energy functions such as the $\text{logcosh}(\cdot)$ prior in a similar manner. Using the approach in the previous section, the partition function can be explicitly expressed as a function of a

$$z(T, \sigma) = \sigma^N z(T, 1) \quad (16)$$

Then using (4) and (16), we can rewrite the normalized log-likelihood of x as

$$\frac{-1}{N} \log \mathcal{P}(x) = \frac{1}{NT} u(x, \sigma) + \log a + \frac{1}{N} \log z(T, 1) \quad (17)$$

Differentiating (17) with respect to a and equating the result to zero yields the ML estimate of σ

$$\left. \frac{\sigma}{N} \frac{d}{d\sigma} u(x, \sigma) \right|_{\sigma=\hat{\sigma}} = -T \quad (18)$$

Due to the non-scalable nature of the partition function, we no longer have a closed form solution for a . However, for a given value of T the solution to (18) can be computed efficiently using 1-D root finding numerical methods.

The derivative of the log of the partition function with respect to T can be computed in a similar manner to $f'(p)$ and is given as

$$\frac{d}{dT} \log z(T, 1) = \frac{1}{T^2} E[u(X, 1) | T, \sigma = 1] \quad (19)$$

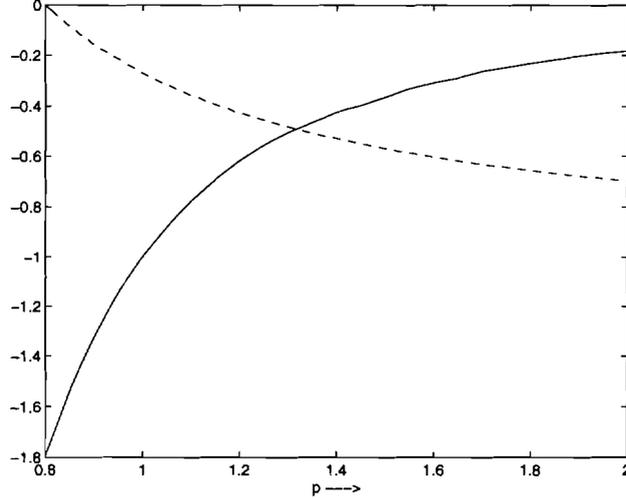


Figure 2: The solid line shows $f'(p)$ and the dashed line shows $f(p)$.

Therefore differentiating (17) with respect to T and equating the result to zero yields the ML estimate of T

$$u(x, \sigma) = E [u(X, 1) | \hat{T}, \sigma = 1] \quad (20)$$

The expectation on the right hand side of equation (20) can be computed prior to the estimation process in a similar manner to $f'(p)$.

The ML estimates of T and a are obtained by solving (18) and (20) simultaneously. However in practice, we have found the simultaneous estimation of T and a to be numerically unstable for the $\text{logcosh}(\cdot)$ prior. Even small noise in the computation of the expectation tends to mask the minima. In practice, T tends to infinity and a tends to zero while their product tends to a constant. For this case, the $\text{logcosh}(\cdot)$ prior tends to the Laplacian distribution. However, the results of (18) and (20) are still useful because if either T or a is known, then the other parameter can be accurately estimated.

4 Model Estimation from Incomplete Data

The ML estimates of the free parameters for the prior image model are a function of the image x as shown in the previous section. But in many cases such as tomographic reconstruction, the image x is unknown. This is an example of the incomplete data problem for which the EM technique [28] was designed. Let θ be the parameter vector that needs to be estimated. Let $\hat{\theta}_k$ be the estimate of θ at the k^{th} iteration. Then a single update of the EM algorithm is given as

$$\hat{\theta}_{k+1} = \arg \max_{\theta} E[\log P_{\theta}(X) | Y = y, \hat{\theta}_k] \quad (21)$$

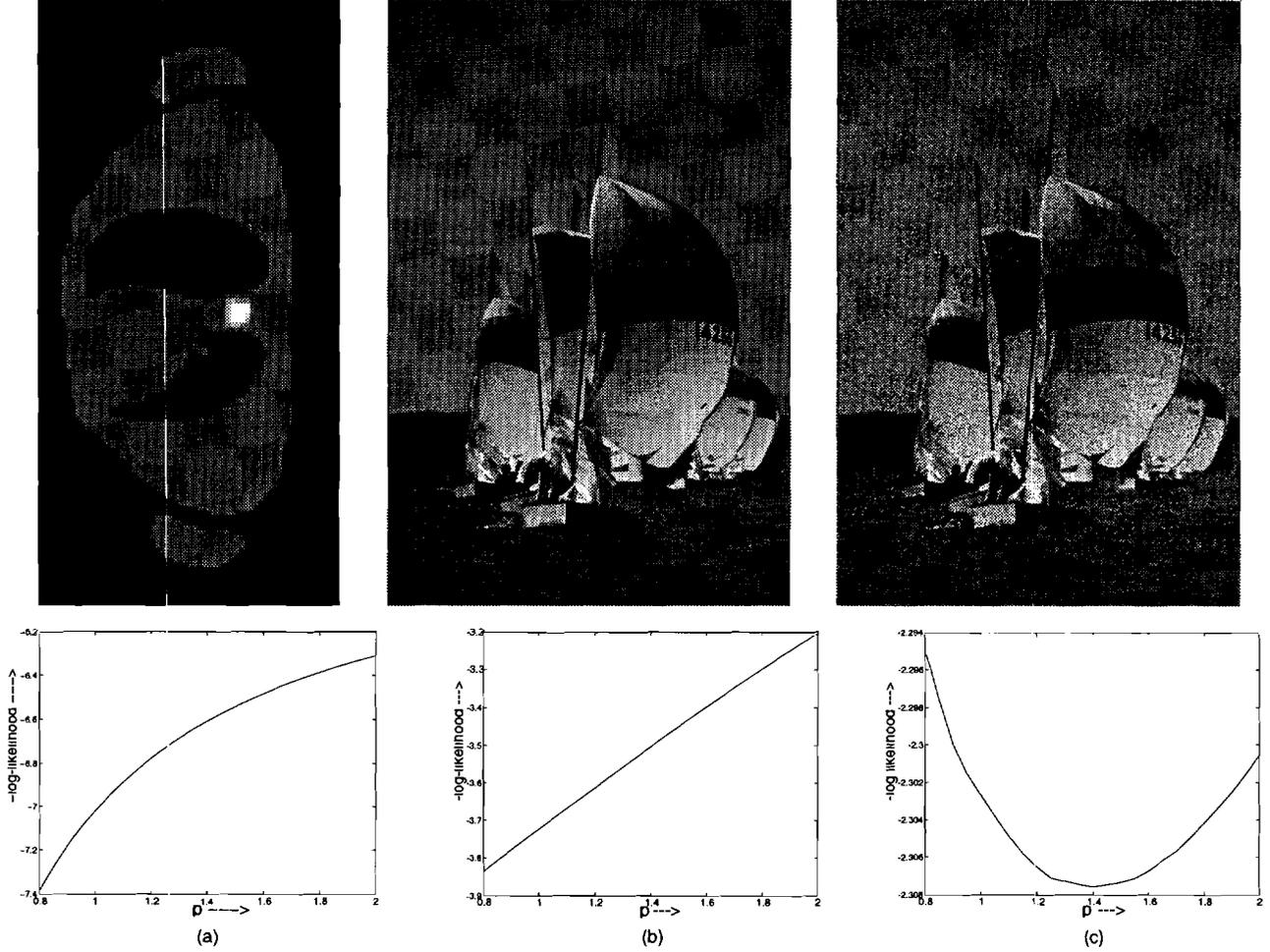


Figure 3: ML estimation of p for (a) transmission phantom (b) natural image (c) image (b) corrupted with Gaussian noise. The plot below each image shows the corresponding negative log-likelihood as a function of p . The ML estimate is the value of p that minimizes the plotted function.

For the GGMRF prior, $\theta = [ap]$. Then using (10) in (21), we obtain the EM update for a and p as

$$(\hat{\sigma}_{k+1}, \hat{p}_{k+1}) = \arg \min_{(\sigma, p)} \left\{ \frac{1}{Np\sigma^p} E[u(X, p) | Y = y, \hat{\sigma}_k, \hat{p}_k] + \log \sigma + \frac{\log p}{p} + f(p) \right\} \quad (22)$$

This can be reduced to a one dimensional optimization by following the procedure of section 3.2. The joint EM updates of p and a are given as

$$\hat{p}_{k+1} = \arg \min \left\{ \log \bar{\sigma}_k(p) + \frac{\log p}{p} + \frac{1}{p} + f(p) \right\} \quad (23)$$

$$\hat{\sigma}_{k+1} = \bar{\sigma}_k(\hat{p}_{k+1}) \quad (24)$$

$$\bar{\sigma}_k(p) \triangleq \left\{ \frac{1}{N} E[u(X, p) | Y = y, \hat{\sigma}_k, \hat{p}_k] \right\}^{1/p}$$

For a logcosh(\cdot) prior, $\theta = [T p]$. The update equations are derived in a manner similar to the GCMRF case. The updated parameters, $[\hat{T}_{k+1}, \hat{\sigma}_{k+1}]$, are obtained by solving the following equations

$$\frac{\sigma}{N} E \left[\frac{d}{d\sigma} u(X, \sigma) | Y = y, \hat{T}_k, \hat{\sigma}_k \right] = -T \quad (25)$$

$$E [u(X, \sigma) | Y = y, \hat{T}_k, \hat{\sigma}_k] = E [u(X, 1) | T, \sigma = 1] \quad (26)$$

simultaneously with respect to T and σ . When T is assumed known, we just solve equation (25) to update the scaling parameter a . On the other hand, if a is assumed known, then equation (26) is solved to update T .

While the variances of emission measurements are proportional to the cross section to be estimated, the variance of transmission data depends on the an external dosage parameter y_T . The data for transmission tomography is sometimes recorded in the form $z_i = \ln(\frac{y_T}{y_i})$. This preserves the ML estimates of integral densities, but results in the loss of the scaling factor for y_T and y_i . Because the variance of the measurements $\{z_i\}$ is approximately inversely proportional to y_T [38], estimation of this parameter is clearly of interest for statistical reconstruction. Using the NIL estimate of y_T derived in the Appendix A and (21), the EM update for y_T is given as

$$\hat{y}_{T_{k+1}} = \frac{M}{2} \left\{ \sum_{i=1}^M E \left[\left\{ e^{-A_{i*} X} + A_{i*} X e^{-z_i} \right\} | Z = z, \hat{\theta}_k, \hat{y}_{T_k} \right] - e^{-z_i} - z_i e^{-z_i} \right\}^{-1} \quad (27)$$

5 Fast Simulation Technique

The EM parameter updates derived in the previous section require the expectation of functions of X . Direct computation of this expectation is intractable, but we can approximate it with spatial averages from realizations of X generated using a stochastic simulation method. The well-known Metropolis algorithm [35] tends to suffer from slow convergence to the equilibrium distribution because the transition probability distribution is required to be symmetric. In this section, we propose a fast simulation technique for efficient computation of the expectation of functions of X .

Hastings [39] and Peskun [40] developed a generalization of the Metropolis algorithm which compensates for asymmetric transition probabilities through the proper choice of the associated acceptance probability. More specifically, 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

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

We note that 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 [41] when the new state for pixel j is generated using the conditional distribution, under $\pi(x)$, of x , given the values of all other pixels. For the Gibbs sampler, $\alpha(x, x') = 1$ and we always accept the new state.

Informal heuristics suggest that a higher acceptance rate would lead to faster convergence, so using the Gibbs sampler to generate the new state would be desirable. Towards this end, let us first examine the form of the conditional distribution of x_j . Let x^n be the image at the n^{th} iteration. Then for the emission case, from (2), (4) and (5), we have

$$\begin{aligned} \log P(x_j | \{X_k = x_k^n : k \neq j\}, y) = \\ \sum_i (-A_{ij}x_j + y_i \log\{A_{ij}(x_j - x_j^n) + A_{i*}x_j^n\} - \frac{1}{T} \sum_{k \in \mathcal{N}_j} b_{j-k} \rho(x_j - x_k^n, \theta) + C \end{aligned} \quad (28)$$

where C is constant independent of x_j and $x_j > 0$. Note that directly generating Samples from (28) would be very computationally expensive. Green and Han [42] suggested using a Gaussian distribution instead with parameters chosen to maximize the acceptance probability. They achieve this by performing a quadratic approximation to the conditional distribution. However, due to the non-quadratic nature of the prior term, the quadratic approximation is good only for the data term [36] in (28). 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 (28)

$$\begin{aligned} \log P(x_j | \{X_k = x_k^n : k \neq j\}, y) \approx \\ d_1(x_j - x_j^n) + \frac{d_2}{2}(x_j - x_j^n)^2 - \frac{1}{T} \sum_{k \in \mathcal{N}_j} b_{j-k} \rho(x_j - x_k^n, \theta) + C' \quad \text{if } x_j > 0 \end{aligned} \quad (29)$$

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 [36] it is shown that

$$\begin{aligned} d_1 &= -\sum_i A_{ij} \left(1 - \frac{y_i}{\tilde{p}_i^n} \right) \\ d_2 &= -\sum_i y_i \left(\frac{A_{ij}}{\tilde{p}_i^n} \right)^2 \end{aligned}$$

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.

$$\begin{aligned} d_1 &= -\sum_i A_{ij} (y_i - y_T e^{-\tilde{p}_i^n}) \\ d_2 &= -\sum_i A_{ij}^2 y_T e^{-\tilde{p}_i^n} \end{aligned}$$

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 j^{th} 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 (29). However generating positive samples from $q_j(x)$ would be computationally intensive when $m \ll 0$. Note that in this case the truncated Gaussian distribution for generating a positive sample appears to be an exponential distribution. In the light of the above discussion, we choose $q_j(x)$ with the following form

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

where m is the mode of the approximated conditional distribution (29)

$$m = \arg \max_x \left[d_1(x - x_j^n) + \frac{d_2}{2}(x - x_j^n)^2 - \frac{1}{T} \sum_{k \in \mathcal{N}_j} b_{j-k} \rho(x - x_k^n, \theta) \right] \quad (30)$$

Choosing s^2 is more difficult due to the prior term. Since we can at best do an approximate fit to the original distribution (28), 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 (29) 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 (28) at $x_j = 0$,

$$\beta = \left\{ -d_1 + d_2 x_j^n + \frac{1}{T} \sum_{k \in \mathcal{N}_j} b_{j-k} \rho'(-x_k^n, \theta) \right\}^{-1}$$

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 (30) 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 [36].

We have to modify this procedure when the expectation of \mathbf{X} with respect to only its prior distribution is required as in the computation of the derivative of the log partition functions used in (15) and (19). For the GGMRF case, the distribution appears to be Gaussian for $p \approx 2$ and Laplacian for $p \approx 1$. Similarly for the logcosh(\cdot) prior case, the distribution appears to be Gaussian for $T \ll 1$ and Laplacian for $T \gg 1$. The transition distribution $q_j(\mathbf{x})$ for these two cases is given as

$$q_j(x) = \begin{cases} \frac{1}{\sqrt{2\pi}s} \exp\left\{-\frac{(x-m)^2}{s^2}\right\} & p \geq 1.5 \text{ or } T \leq 1, x > 0 \\ \frac{1}{2\beta} \exp\left\{-\frac{|x-m|}{\beta}\right\} & p < 1.5 \text{ or } T > 1, x > 0 \\ 0 & x < 0 \end{cases}$$

The mode m of the transition distribution is again chosen as the mode of the prior distribution and is given as

$$m = \arg \min_x \left[\frac{1}{T} \sum_{k \in \mathcal{N}_j} b_{j-k} \rho(x - x_k^n, \theta) \right]$$

The parameters s^2 and β are difficult to choose in this case. One alternative is to use the least squares value of the parameter by matching the derivatives of the prior distribution and the fit at a few points in the neighborhood of m . In our experiments however, we noticed that the simulations are accurate if we hand pick a large value for the variances of the fitted distributions. In particular, we chose $s^2 = 50$ and $\beta = 50$ to compute $f'(p)$ in section 3.2.

6 Extrapolation of Parameter Estimates

In this section, we propose a method to extrapolate the parameter estimates when the simulation is terminated prematurely prior to convergence. We achieve this by doing a piecewise linear approximation to the gradient of the log-likelihood of the data.

For the GGMRF case, we will derive the gradients in terms of T and p since it results in simpler expressions. The substitution $T = p\sigma^p$ can be used to revert back to the scale parameter σ . Using the result of Appendix B, the normalized gradient of the log-likelihood with respect to T is obtained as

$$g(T) \triangleq \frac{pT^2}{N} \frac{\partial}{\partial T} \log P(y) = \frac{p}{N} E[u(X, p) | Y = y, T, p] - T \quad (31)$$

The ML estimate of T by definition is obtained as the root of $g(T)$

$$\frac{p}{N} E[u(X, p) | Y = y, \hat{T}_{ML}, p] - \hat{T}_{ML} = 0$$

It now becomes clear that the EM algorithm is iteratively trying to solve for the ML estimate by setting

$$\hat{T}_{k+1} = \frac{p}{N} E[u(X, p) | Y = y, \hat{T}_k, p]$$

This is exactly the EM update for σ (24) when $T = p\sigma^p$.

Our goal is to model the normalized gradient $g(T)$ to obtain a better estimate of its root. Note that we can numerically compute $g(T)$ at the current estimate of the parameters when performing an EM update. We then use the numerical value of $g(T)$ computed at n points in the past to obtain a least squares (LS) fit to a line. The root of the LS fit is then the extrapolated value of T . Fig. 4 shows the normalized gradient $g(T)$ with respect to T for an emission phantom when we use only one sample of X to estimate the expectation of $u(X, p)$. Here we use 3 points from the immediate past to obtain the LS fit. The zero crossing of the LS fit yields the extrapolated value of T . Note that the extrapolated value of T is close to the ML estimate after just 4 iterations.

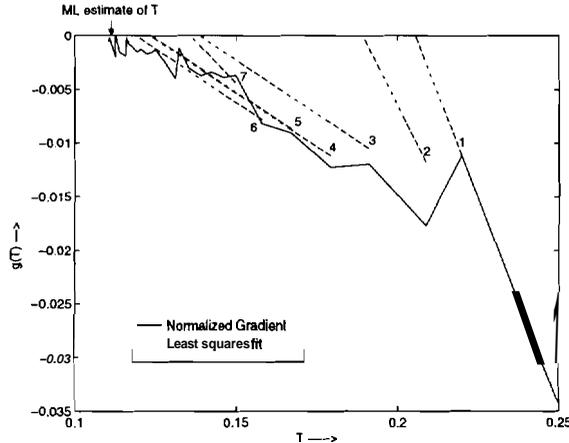


Figure 4: The plot shows the normalized gradient $g(T)$ computed at the EM updates of T for the emission phantom using a GGMRF prior with $p = 1.1$. The least squares (LS) fit obtained at the first 7 points are numbered (1-7) and shown with a dashed line. Three points from the immediate past are used for the LS fit. The root of the LS fit is the extrapolated value of T .

The generalization to the case when p is not known is conceptually easy. Using the result of Appendix B, the normalized gradient of the log-likelihood with respect to p is obtained as

$$\frac{p^2 T}{N} \frac{\partial}{\partial p} \log P(y) = \frac{-p^2}{N} E \left[\frac{d}{dp} u(X, p) | Y = y, T, p \right] + T \log T - p^2 T f'(p) \quad (32)$$

Equations (31) and (32) need to be rooted simultaneously to obtain the ML estimate of T and p . In the same spirit as the one dimensional case, we now use the past n points to obtain a least squares fit to a plane for the gradients. The value of T and p that roots the fitted planes are then the extrapolated value of the parameters.

The logcosh(\cdot) prior case is handled in a similar fashion. In this case, the normalized gradient of the log-likelihood with respect to T and a is given as

$$\frac{\sigma T}{N} \frac{\partial}{\partial a} \log P(y) = \frac{-\sigma}{N} E \left[\frac{d}{d\sigma} u(X, a) \mid Y = y, T, a \right] - T \quad (33)$$

$$\frac{T^2}{N} \frac{\partial}{\partial T} \log P(y) = \frac{1}{N} E [u(X, a) \mid Y = y, T, a] - \frac{1}{N} E [u(X, 1) \mid T, a = 1] \quad (34)$$

Equations (33) and (34) can be rooted simultaneously or individually to obtain the extrapolated values of T and a .

Experimental Results

In this section, we perform simulations on synthetic transmission and emission phantoms to assess the effectiveness of the parameter estimation schemes proposed in this report. The estimated parameters will be used to reconstruct the phantoms in order to show that the ML estimates indeed yield superior quality images.

The synthetic transmission phantom used for the simulations is shown in Fig. 5. The attenuation map is a 128 by 64 array of 4.5mm pixels; it represents a human thorax with linear attenuation coefficients 0.0165/mm, 0.0096/mm, and 0.0025/mm, for bone, soft tissue, and lungs respectively. Poisson random variables were generated from 192 projections taken at 256 uniformly spaced angles to obtain the noisy projection data. The A_{ij} factors correspond to 6mm wide strip integrals on 3mm center-to-center spacing. The dosage parameter y_T was different for each projection and was generated using log-normal variates with a standard deviation of 0.3 to account for the detector efficiency variations. The total photon count was approximately 3 million. The CBP image was obtained by using a generalized hamming filter and is shown in Fig. 5.

The synthetic emission phantom and the corresponding CBP image are shown in Fig. 6. The emission rates are on an array of 128 by 128 1.56mm pixels. Readings were taken at 128 equally spaced angles, and 128 perfectly collimated detectors at each angle. Emission rates were scaled to yield a total count of approximately 3 million.

An 8-point neighborhood system with periodic boundary conditions was used for the simulations. The weights b_{i-j} for each pixel i were normalized to 1 by setting $b_{i-j} = (2\sqrt{2} + 4)^{-1}$ for nearest neighbors and $b_{i-j} = (4\sqrt{2} + 4)^{-1}$ for diagonal neighbors.

A single sample generated from the posterior distribution was used to compute the expectation in the EM update. We verified experimentally that the number of iterations used to compute the expectation had little or no effect on the value of the parameter at convergence. Figure 7(a) shows this result. The initial parameter values for the EM algorithm were computed from the CBP image. The CBP image was also used to initialize the simulation method and the MAP reconstruction.

We noted in section 3 that the optimal value of p for the GGMRF prior for the transmission phantom and all the natural images is less than 1. Therefore in all our simulations, we will consider the simple case of estimating just the scaling parameter a and assume $p = 1.1$ since this results in a convex cost function. We will refer to the fast simulation method proposed in section 5 as the accelerated Metropolis or AM method and the conventional Metropolis method as the CM method. Fig. 7(a) shows the estimation of a using the GGMRF prior for the emission phantom with the CM method. The transition distribution is chosen to be Gaussian with the variance as the free parameter. From the plots we observe that the choice of this variance, s , is crucial to the rate of convergence of the CM method. In practice however, it is unclear how to choose this variance before performing the simulations.

Fig. 7(b) compares the rates of convergence between the CM method and the AM method. The variance of the transition distribution for the CM method is chosen to yield the fastest convergence. Even then, the AM method is seen to converge much faster than the best case CM method. The least squares fit for the extrapolation method was obtained by using the gradient computed at 5 past values of a . It is seen that the extrapolation method yields estimates close to the converged value after just 5 iterations. Fig. 8 shows the corresponding plot for the $\text{logcosh}(\cdot)$ prior with $T = 10$. We will show later on that this value of T for the $\text{logcosh}(\cdot)$ yields reconstructions similar to that of a GGMRF prior with $p = 1.1$.

Fig. 9 shows the convergence plots for estimating a for the transmission phantom using the GGMRF prior with $p = 1.1$ and $p = 2.0$. We observe that the initial value of a computed from the CBP images is very close to $\hat{\sigma}_{ML}$. Hence the AM and CM methods are comparable in this case.

In order to evaluate the quality of the reconstruction obtained using the ML estimate of a , we show in Fig. 10 several reconstructions of the transmission phantom using the GGMRF prior with different values of a . Note that in this case $\hat{\sigma}_{ML} \approx \hat{\sigma}_{CBP}$ and the reconstructions obtained using $\hat{\sigma}_{ML}$ and $\hat{\sigma}_{CBP}$ are similar. Two other reconstructions are shown with a equal to $2\hat{\sigma}_{ML}$ and $\frac{\hat{\sigma}_{ML}}{2}$. Fig. 11 shows the corresponding reconstructions for the emission case. In this case, $\hat{\sigma}_{ML}$ is different from $\hat{\sigma}_{CBP}$ and therefore we show the reconstruction corresponding to $\hat{\sigma}_{CBP}$ also.

For the $\text{logcosh}(\cdot)$ prior, we experimented with different values of T , using the optimal

a in each case for the reconstructions. Fig. 12 and 13 shows the reconstructions for the transmission and emission case respectively. When $T = 1$, the $\text{logcosh}(\cdot)$ prior is similar to the Huber function which applies quadratic cost to small inter pixel difference and linear cost otherwise. Consequently the reconstruction has some smooth variation and the edges are sharp. When $T = 10$, the reconstruction is similar to the GGMRF case with $p = 1.1$. For $T = 100$, the $\text{logcosh}(\cdot)$ prior looks almost Laplacian (i.e. the GGMRF with $p = 1$) and the reconstruction has a tendency to get stuck on edges of the cost function [43]. In this case, the MAP estimate is difficult to compute due to the extremely slow convergence.

8 Conclusion

We have shown in this report that maximum-likelihood estimation of free parameters for unsupervised Bayesian image reconstruction is feasible for a broad selection of image models and problem settings. We have presented an efficient scheme for optimal (estimation of the parameters associated with a general MRF model directly from image samples. By estimating the shape parameter p of a GGMRF, we have shown that natural images appear more Laplacian than Gaussian. Using the accelerated Metropolis algorithm for EM in estimation from incomplete data, coupled with the extrapolation method, we can compute the ML estimates in a few iterations of EM. In fact, we observed that in cases where we have a high signal to noise ratio, the estimate obtained from the CBP image is very close to the optimal value and there is no need to perform the EM updates.

While only the GGMRF and the $\text{logcosh}(\cdot)$ prior models were used for the reconstructions presented here, the proposed method can easily be adapted to any other model. We observed that the $\text{logcosh}(\cdot)$ with a large value of T yields reconstructions comparable in quality to the GGMRF prior for $p \approx 1$. However, the parameter estimation for the GGMRF prior is much simpler and computationally less expensive than for the $\text{logcosh}(\cdot)$ prior. This advantage, which may be substantial in unsupervised reconstruction, stems from the scalable nature of the energy function of the GGMRF prior.

Appendix A

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 $\{Z_i = \ln(\frac{y_T}{Y_i})\}$. Consider the log-likelihood function of z in terms of the unknown dosage parameter y_T . Let the i -th actual discretized projection measurement across X be $\tilde{z}_i = A_{i*}x$. Note that Y_i are Poisson

distributed with mean and variance $y_T e^{-z_i}$. Then by a simple transformation, we have

$$\mathcal{P}(Z_i = z_i | X = x) = \frac{\exp\{-y_T e^{-z_i}\} (y_T e^{-z_i})^{y_T e^{-z_i}}}{(y_T e^{-z_i})!} \quad (35)$$

for values of z_i corresponding to positive integer values of y_i . Stirling's formula provides a simplifying approximation for the factorial, which is relatively accurate for numbers in the typical range of transmission photon counts [44]:

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

Using this substitution, differentiating the logarithm of (35) with respect to y_T and setting the result to zero yields the ML estimate of y_T .

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

Appendix B

In this appendix we derive the gradient of the log-likelihood of the observations \mathbf{y} with respect to the prior model parameters. Let θ be the parameter vector. Consider the log-likelihood of \mathbf{y}

$$\log P_\theta(\mathbf{y}) = \log P_\theta(\mathbf{x}, \mathbf{y}) - \log P_\theta(\mathbf{x} | \mathbf{y})$$

Taking the expectation conditioned on \mathbf{Y} and $\theta = \theta'$ of the above equation., we have

$$\log P_\theta(\mathbf{y}) = Q(\theta; \theta') - H(\theta; \theta') \quad (36)$$

where

$$Q(\theta; \theta') = E[\log P_\theta(\mathbf{X}, \mathbf{y}) | Y = \mathbf{y}, \theta'] \quad (37)$$

$$H(\theta; \theta') = E[\log P_\theta(\mathbf{X} | \mathbf{y}) | Y = \mathbf{y}, \theta'] \quad (38)$$

Differentiating (36) with respect to θ and setting $\theta' = \theta$, we obtain

$$\frac{d}{d\theta} \log P_\theta(\mathbf{y}) = Q^{10}(\theta; \theta) - H^{10}(\theta; \theta) \quad (39)$$

where the superscript ¹⁰ denotes the derivative of the function with respect to its first argument. Note that since

$$\begin{aligned} H^{10}(\theta; \theta) &= \int_x \log P_\theta(x | \mathbf{y}) \frac{d}{d\theta} \log P_\theta(x | \mathbf{y}) dx \\ &= 0 \end{aligned} \quad (40)$$

we have

$$\frac{d}{d\theta} \log P_{\theta}(y) = Q^{10}(\theta; 8) \quad (41)$$

Substituting (37) in (41) and using the fact that X is sufficient for θ , we obtain

$$\frac{d}{d\theta} \log P_{\theta}(y) = E\left[\frac{d}{d\theta} \log P_{\theta}(X) | Y = y, \theta\right] \quad (42)$$

Acknowledgment

The authors would like to thank prof. J. Fessler of the university of Michigan for providing the transmission phantom.

References

- [1] S. Geman and D. McClure, "Bayesian images analysis: An application to single photon emission tomography," Proc. Statist. *Comput.* sect. Amer. Stat. Assoc., 1985, Washington, DC, pp. 12–18.
- [2] B. Hunt, "Bayesian methods in nonlinear digital image restoration," *IEEE Trans. on Comput.*, vol. c-26, no. 3, pp. 219–229, 1977.
- [3] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Int. Stat. Inst.*, vol. LII-4, pp. 5–21, 1987.
- [4] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-d Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. on Medical Imaging*, vol. 8, no. 2, pp. 194–202, June 1989.
- [5] P. J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm," *IEEE Trans. on Medical Imaging*, vol. 9, no. 1, pp. 84–93, March 1990.
- [6] K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. on Signal Processing*, vol. 41, no. 2, February 1993.
- [7] J. Fessler, "Hybrid poisson/polynomial objective functions for tomographic image reconstruction from transmission scans," *IEEE Trans. on Image Processing*, vol. 4, no. 10, pp. 1439–1450, October 1995.
- [8] E. Levitan and G. Herman, "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. on Medical Imaging*, vol. MI-6, pp. 185–192, Sept. 1987.
- [9] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, Massachusetts: MIT Press., 1987.
- [10] R. Stevenson and E. Delp, "Fitting curves with discontinuities," Proc. of the first international workshop on robust computer vision, pp. 127–136, October 1-3 1990.
- [11] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE' Trans. on Medical Imaging*, vol. 9, no. 4, pp. 439–446, December 1990.

- [12] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, 1992.
- [13] C. A. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving map estimation," *IEEE Trans. on Image Processing*, vol. 2, pp. 296–310, July 1993.
- [14] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*. New York: Springer-Verlag, 1988.
- [15] B. Hunt, "The application of constrained least squares estimation to image restoration by digital computer," *IEEE Trans. on Comput.*, vol. c-22, no. 9, pp. 805–812, September 1973.
- [16] P. Hall and D. Titterton, "Common structure of techniques for choosing smoothing parameters in regression problems," *Journal of the Royal Statistical Society B*, vol. 49, pp. 184–198, 1987.
- [17] J. Rice, "Choice of smoothing parameter in deconvolution problems," in *Contemporary Mathematics* (J. S. Maron, ed.), vol. 59, pp. 137–151, Providence, RI: American Math. Soc., 1986.
- [18] S. Reeves and R. Mersereau, "Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration," *Optical Engineering*, vol. 29, no. 5, pp. 446–454, May 1990.
- [19] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Identification and restoration of noisy blurred images using the expectation-maximization algorithm," *IEEE Trans. on Acoust. Speech and Signal Proc.*, vol. 38, no. 7, pp. 1180–1191, July 1990.
- [20] A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterton, "A study of methods of choosing the smoothing parameter in image restoration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 326–339, 1991.
- [21] 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. on Image Processing*, vol. 1, no. 3, pp. 322–336, July 1992.
- [22] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society B*, vol. 36, no. 2, pp. 192–236, 1974.
- [23] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society B*, vol. 48, no. 3, pp. 259–302, 1986.
- [24] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. FAMI-9, no. 1, pp. 39–55, January 1987.
- [25] S. Geman and C. Graffigne, "Markov random field image models and their applications to computer vision," *Proc. of the Intl Congress of Mathematicians*, 1986, Berkeley, California, pp. 1496–1517.
- [26] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 8, pp. 799–813, August 1989.

- [27] A. Gray, J. Kay, and D. Titterton, "An empirical study of the simulation of various models used for images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 507–513, May 1994.
- [28] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 39, no. 1, pp. 1–38, 1977.
- [29] J. Zhang, "The mean field theory in EM procedures for blind Markov random field image restoration," *IEEE Trans. on Image Processing*, vol. 2, no. 1, pp. 27–40, January 1993.
- [30] R. Schultz, R. Stevenson, and A. Lumsdaine, "Maximum likelihood parameter estimation for non-Gaussian prior signal models," *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. 2, November 1994, Austin, TX, pp. 700–704.
- [31] W. Pun and B. Jeffs, "Data directed MAP image restoration using an adaptive generalized Gauss-Markov random field model," (submitted to) *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, May 7-10 1996, Atlanta, GA.
- [32] C. A. Bouman and K. Sauer, "Maximum likelihood scale estimation for a class of Markov random fields," *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, vol. 5, April 19-22 1994, Adelaide, South Australia, pp. 537–540.
- [33] K. Lange, "An overview of Bayesian methods in image reconstruction," *Proc. of the SPIE Conference on Digital Image Synthesis and Inverse Optics*, vol. SPIE-1351, 1990, San Diego, CA, pp. 270–287.
- [34] Y. Ogata, "A Monte Carlo method for an objective Bayesian procedure," *Ann. Inst. of Statist. Math.*, vol. 42, no. 3, pp. 403–433, 1990.
- [35] 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.
- [36] C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, to appear March 1996.
- [37] P. Huber, *Robust Statistics*. New York: John Wiley & Sons, 1981.
- [38] G. Herman, *Image Reconstruction from Projections: The Fundamental of Computerized Tomography*. New York: Academic Press, 1980.
- [39] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
- [40] P. H. Peskun, "Optimum Monte-Carlo sampling using Markov chains," *Biometrika*, vol. 60, no. 3, pp. 607–612, 1973.
- [41] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 721–741, Nov. 1984.
- [42] 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. Piccioni, eds.), pp. 142–164, Berlin: Springer-Verlag, 1992.

- [43] K. Sauer and C. Bouman, "Bayesian estimation of transmission tomograms using segmentation based optimization," *IEEE Trans. on Nuclear Science*, vol. 39, pp. 1144–1152, 1992.
- [44] M. Abramowitz and I. Stegun, eds., *Handbook of Mathematical Functions*. New York: Dover, 1965.

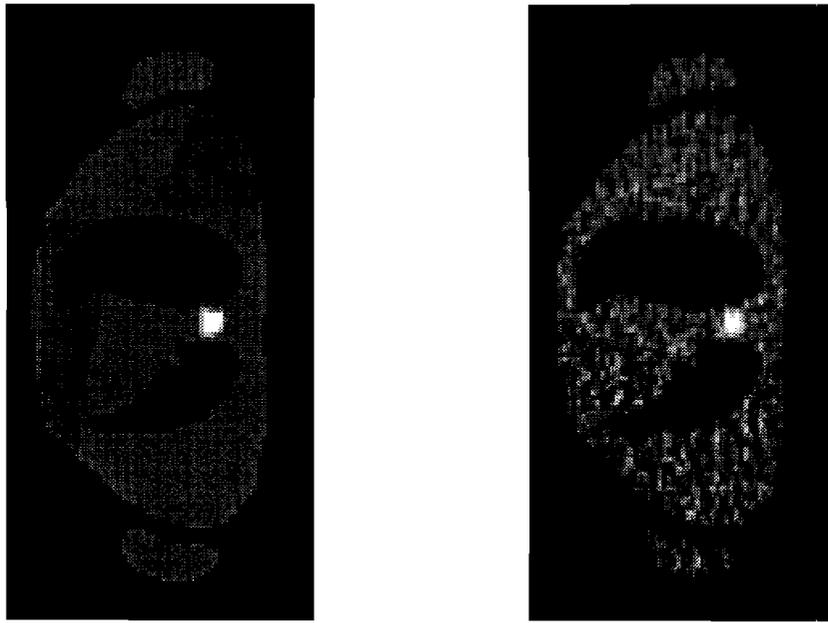
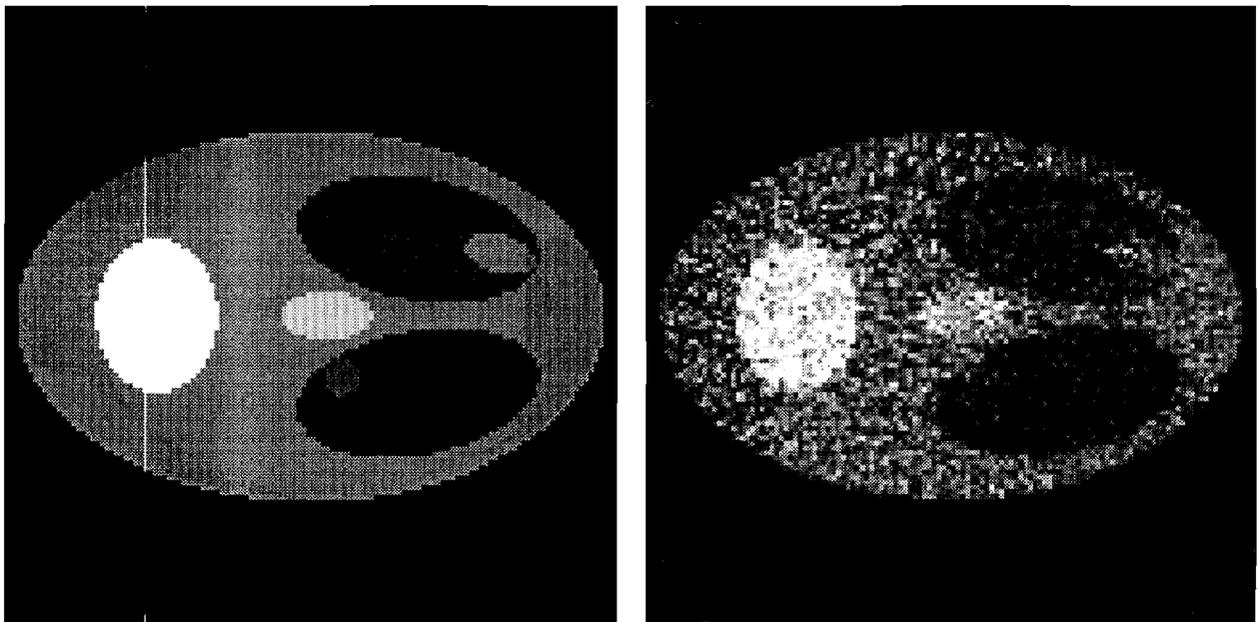


Figure 5: Original transmission phantom and CBP reconstruction. (Phantom courtesy of J. Fessler, university of Michigan)



a	b
---	---

Figure 6: (a) Original emission phantom and (b) CBP reconstruction.

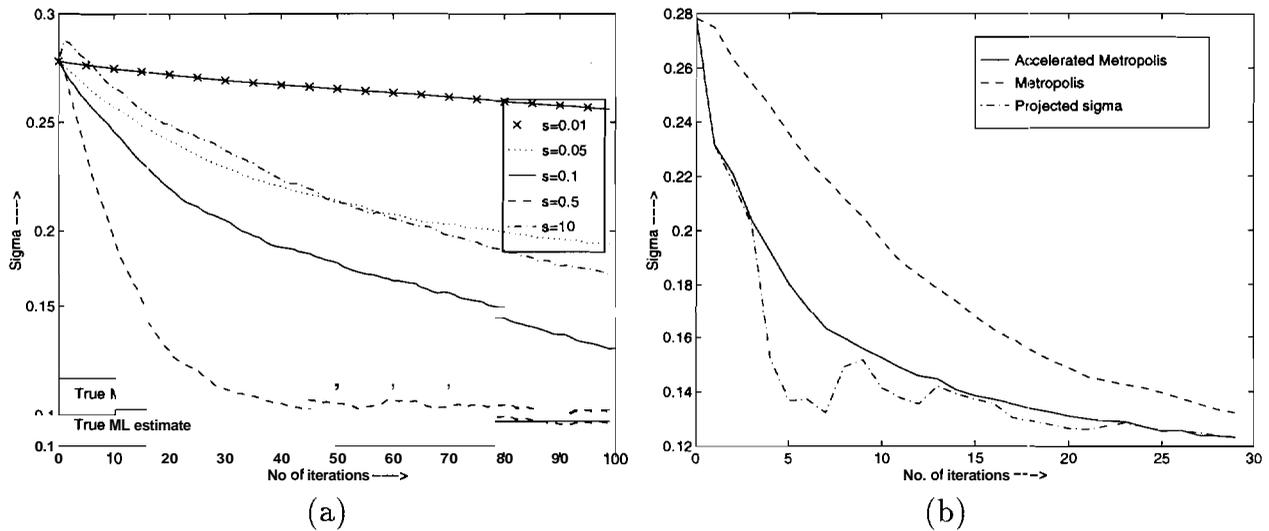


Figure 7: The above plots show the EM updates for a for the emission phantom modeled by a GGMRF prior ($p = 1.1$) using (a) conventional Metropolis (CM) method, (b) accelerated Metropolis (AM) and the extrapolation method. The parameter s denotes the standard deviation of the symmetric transition distribution for the CM method. All the updates are done using a single sample of \mathbf{X} to compute the expectation. The true NIL estimate is the converged value of a when 50 samples are used to compute the expectation.

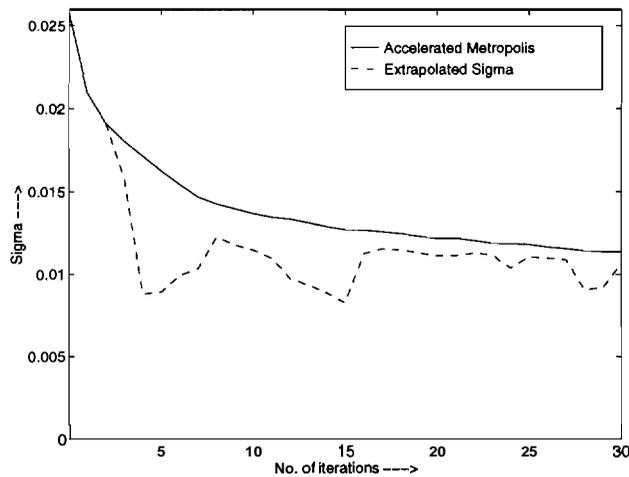


Figure 8: The above plots shows the EM updates for a using the accelerated Metropolis method and the extrapolated value of a for the emission phantom using the $\text{logcosh}(\cdot)$ prior with $T = 10$.

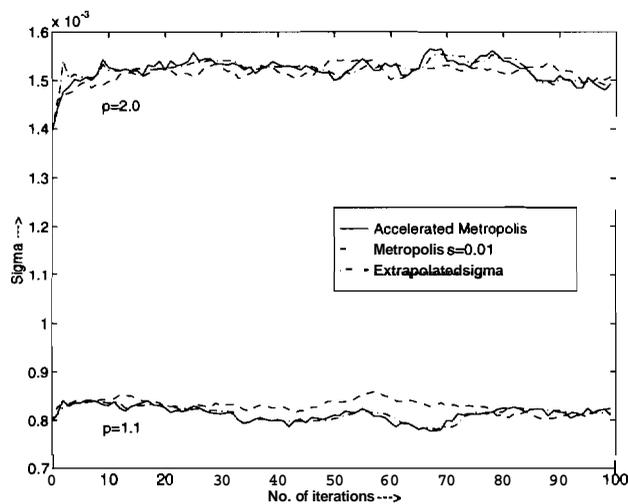


Figure 9: The above plots shows the EM updates for a using the Metropolis method, accelerated Metropolis method, and the extrapolated value of a for the transmission phantom using the GGMRF prior.

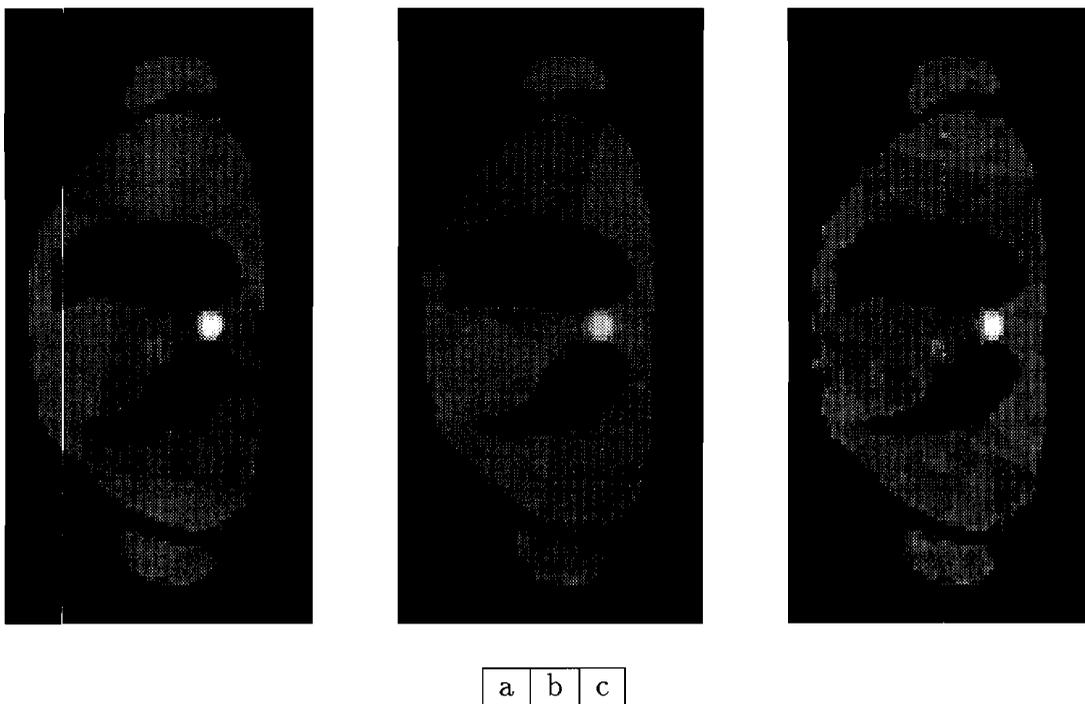
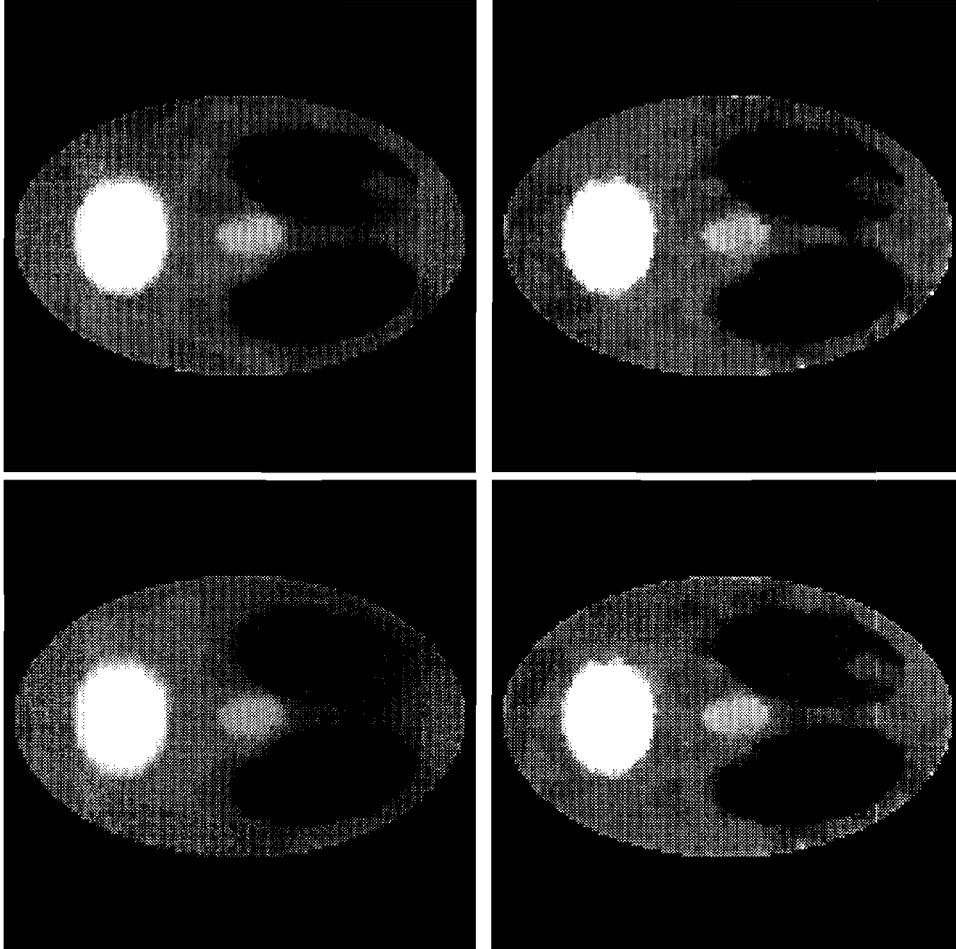


Figure 10: Reconstructed transmission phantom using GGMRF prior with $p = 1.1$ The scale parameter a is (a) $\hat{\sigma}_{ML} \approx \hat{\sigma}_{CBP}$, (b) $\frac{1}{2}\hat{\sigma}_{ML}$, and (c) $2\hat{\sigma}_{ML}$



a	b
c	d

Figure 11: Reconstructed emission phantom using GGMRF prior with $p = 1.1$. The scale parameter σ is (a) $\hat{\sigma}_{ML}$, (b) $\hat{\sigma}_{CBP}$, (c) $\frac{1}{2}\hat{\sigma}_{ML}$, and (d) $2\hat{\sigma}_{ML}$.

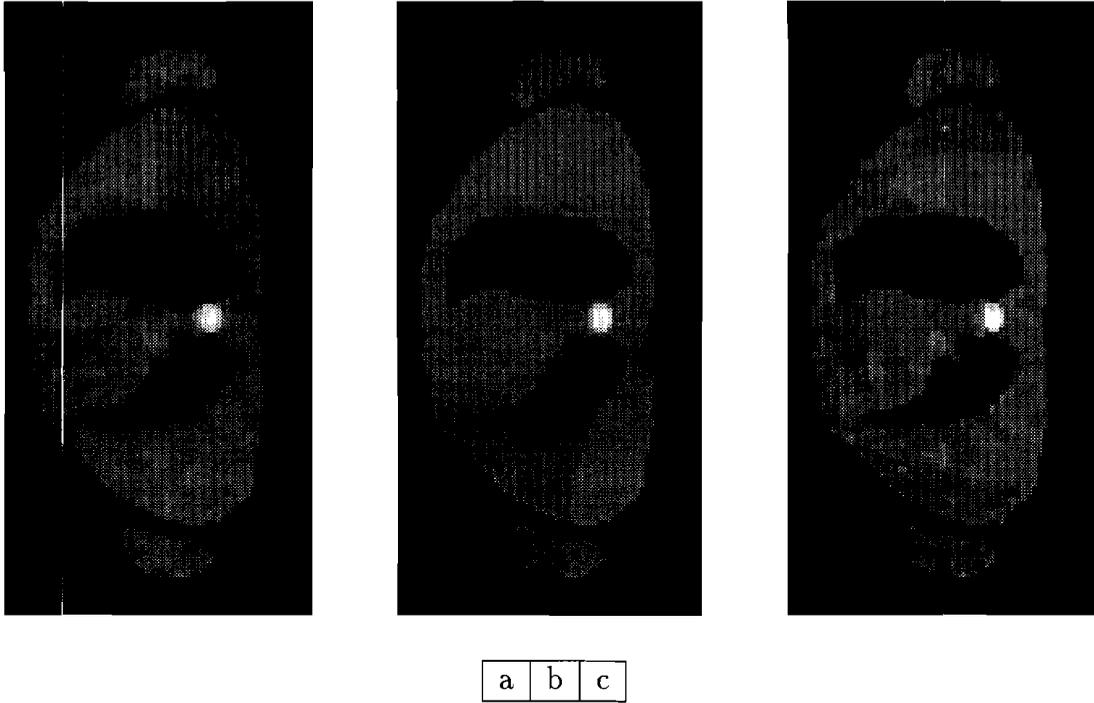


Figure 12: Reconstructed transmission phantom using $\text{logcosh}(\cdot)$ prior with the scale parameter a optimally estimated for different values of T . The value of T is (a) 1, (b) 10, and (c) 100.

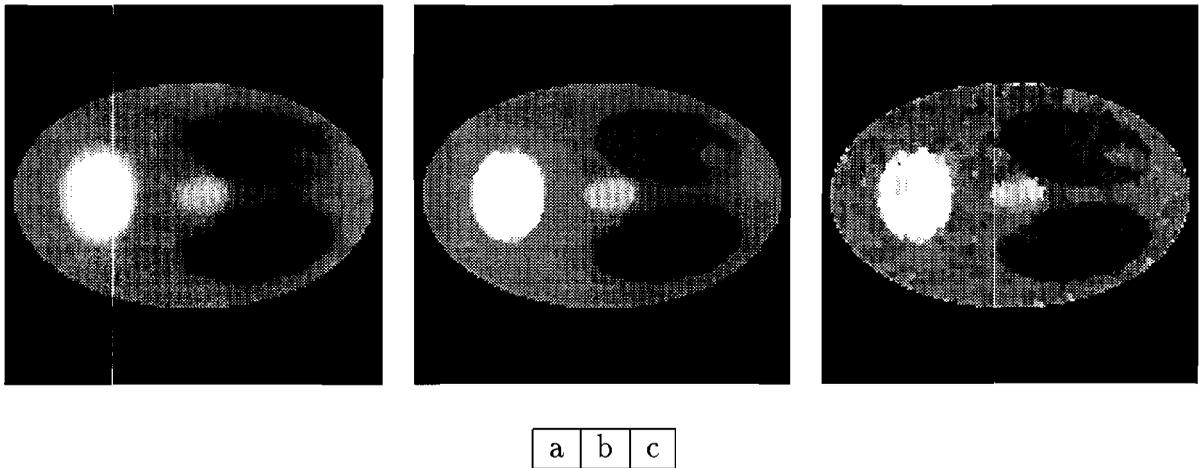


Figure 13: Reconstructed emission phantom using $\text{logcosh}(\cdot)$ prior with the scale parameter a optimally estimated for different values of T . The value of T is (a) 1, (b) 10, and (c) 100.