TRACTABLE MODELS AND EFFICIENT ALGORITHMS FOR BAYESIAN TOMOGRAPHY

Charles A. Bouman[1], *Ken Sauer*[2] *and Suhail S. Saquib*[1]

[1]School of Electrical Engineering, Purdue University, West Lafayette, IN 47907.
[2]Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556.

ABSTRACT

Bayesian methods have proven to be powerful tools for computed tomographic reconstruction in realistic physical problems. However, Bayesian methods require that a number of modeling and computational problems be addressed. This paper summarizes a coherent system of statistical modeling and optimization techniques designed to facilitate efficient, unsupervised Bayesian emission and transmission tomographic reconstruction. New results are included on improved convergence behavior of these methods.

1. INTRODUCTION

For the past 10 years, Bayesian estimation techniques have been studied for application in computed tomography, providing a framework for accurately modeling physical measurements and incorporating prior knowledge[1, 2, 3]. These model based techniques can improve the quality of reconstructions by incorporating more information about the data and measurement process. In particular, Bayesian reconstruction methods commonly employ a forward model which accounts for effects such as of photon counting statistics, missing or attenuated projections, and alternative projection geometries. Reconstruction quality can also be substantially improved by using a prior image model which includes positivity constraints and non-Gaussian image statistics to preserve edges and reduce noise. The advantages of this Bayesian approach are particularly important when the signal-to-noise ratio is low, the dynamic range of material densities is high, or the projection measurements are sparse or attenuated. In contrast, the quality of more conventional direct reconstruction approaches such as filtered back projection (FBP) can be severely limited in these cases.

While Bayesian reconstruction methods have numerous potential advantages, important technical barriers still remain to wide spread application. First, Bayesian reconstruction methods usually require iterative numerical optimization of a cost function with respect to all the image pixels. Because of the high dimensionality, this optimization can be very computationally intensive, often requiring orders of magnitude more computation than conventional FBP reconstructions[4]. The complexity of this optimization is worsened by the nonlinearities introduced in accurate modeling. Second, the forward and prior models for image reconstruction usually contain unknown free parameters which dramatically effect the quality of reconstructions. The most important among these parameters is the so called "regularization constant" which is related to the signal-to-noise ratio of the data.

The objective of this paper is to present a coherent system of models and algorithms for Bayesian tomography which accurately represent the data collection process and prior knowledge, estimate unknown model parameters directly from the available data, and substantially reduce the cost of the required numerical operations. We present new results illustrating substantial computational savings in finding both MAP reconstructions and estimates of the parameters which determine their character.

2. MODELING

The standard forward model which we use assumes Poisson distributed photon counting statistic for the tomographic measurements. The two general cases which we consider are the transmission model used for measuring attenuation cross-sections, and the emission model used to measure the spatial density of photon emissions. The emission model may be applied to a wide variety of applications including single photon emission tomography (SPECT), positron emission tomography (PET), and photon limited imaging.

In the Bayesian framework, the image cross-section is also modeled with an *a priori* distribution. We choose a Markov random field (MRF) model with a distribution of the form

$$p(x) = \begin{cases} \frac{1}{z}\exp\left\{\sum_{\{i,j\}} a_{i,j}\rho\left(\frac{x_i - x_j}{\sigma}\right)\right\} & \text{if } \forall i,\ x_i \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\rho(\cdot)$ penalizes differences between adjacent pixels, and σ is a parameter which controls the variation in x. A variety of convex functions have been proposed for $\rho(\cdot)$ [5, 6, 7]. In this work, we use a Generalized Gaussian Markov random field (GGMRF) [7] where $\rho(x) = \frac{1}{q}|x|^q$ for $p \geq 1$. This yields the following distribution.

$$p(x) = \begin{cases} \frac{1}{z}\exp\left\{\sum_{\{i,j\}} a_{i,j}\frac{1}{q\sigma^q}|x_i - x_j|^q\right\} & \text{if } \forall i,\ x_i \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The GGMRF model is convex, allows preservation of edges, leads to tractable methods for estimating the parameter σ, and does not require the choice of a threshold parameter related to edge magnitude.

THIS WORK WAS SUPPORTED BY NATIONAL SCIENCE FOUNDATION GRANT NUMBER MIP93-00560.

0-7803-2431-5/95 $4.00 © 1995 IEEE

3. OPTIMIZATION

While the mathematical forms of the transmission and emission models are different, they have a similar structure which allows them to be solved using the same algorithmic approaches[8, 9]. The transmission tomographic reconstruction problem under various forward models has been approached with many different types of numerical tools. However, since the seminal paper of Shepp and Vardi[10], the principal method for statistical emission tomographic reconstruction has been the expectation-maximization (EM) algorithm[11]. Although improvements in convergence have been made to EM for tomographic reconstruction[12], it is still slow to converge due to its similarity to gradient ascent[13]. Application of EM to the Bayesian maximum *a posteriori* (MAP) problem is non-trivial, since the maximization step is complicated by the local interaction of image pixels in most prior distributions. Several techniques have been employed to overcome this limitation[5, 14, 15], but the problem of the slow convergence of the basic EM algorithm for this problem remains.

The key to our approach is an efficient algorithm for minimizing the Bayesian cost function with respect to a single pixel's value[9, 13]. This local update strategy has been widely used in optimization problems and goes by a variety of names including Gauss-Seidel (GS) in partial differential equations, iterative coordinate decent (ICD) in numerical optimization, and iterated conditional modes (ICM) used in maximum a posteriori (MAP) estimation. A similar method known as the method of space-alternating generalized expectation-maximization (SAGE)[16, 17] has been proposed by Fessler and Hero. For the emission problem, the SAGE method is similar in operation and performance to ICD since it incorporates a sequential update into EM by varying the complete data space.

For a variety of reasons, the ICD method is particularly well suited to the problem of Bayesian tomography. First, ICD can be efficiently applied to the log likelihood expressions resulting from accurate forward models of Poisson counting statistics. Second, the ICD algorithm converges very rapidly when initialized with the filtered back projection (FBP) reconstruction. The fast numerical convergence of ICD results from the fact that the tomographic reconstruction problem is the solution to an integral equation. This structure causes the high spatial frequencies to converge rapidly when pixels are updated using the ICD algorithm.

The third important advantage of the ICD algorithm is that it easily incorporates convex constraints, and non-Gaussian prior distributions. In particular, positivity is an important convex constraint which can both improve the quality of reconstructions and significantly speed numerical convergence. Positivity constraints are particularly important in emission reconstruction problems where typically a great deal of the image cross-section is zero. Non-Gaussian prior distributions such as the GGMRF are also important because they can substantially reduce noise while preserving edge detail.

In Figures 1 and 2, rates of convergence are compared in computing MAP emission tomographic reconstructions using ICD and several adaptations of the EM algorithm to the Bayesian problem. The data include approximately 5×10^4 counts among 64×64 projections. The alternatives shown are Green's one-step late (OSL)[5], a generalized expectation-maximization (GEM) algorithm of Hebert and Leahy[14] and De Pierro's method[15]. The ICD iterations proceed to the MAP estimate at a much more rapid pace than the EM-based techniques, showing promise for much more computationally efficient estimation.

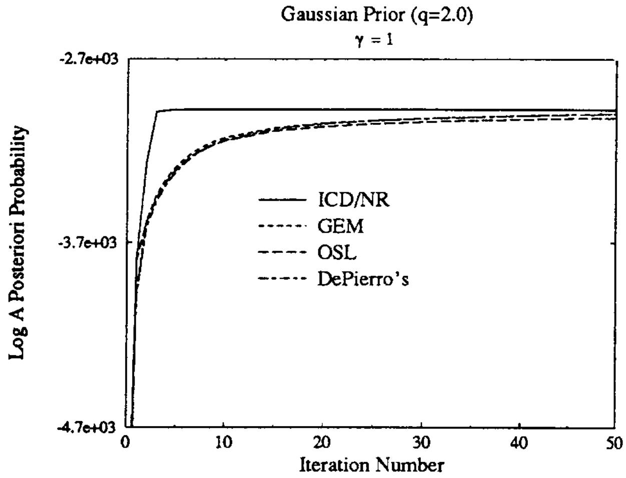

Figure 1: Convergence of MAP estimates using ICD/Newton-Raphson updates, Green's (OSL), Hebert/Leahy's GEM, and De Pierro's method, with a Gaussian prior model.

4. PARAMETER ESTIMATION

Parameter estimation for Markov random fields is known to be a difficult problem due to the intractable form of the normalizing constant or partition function. However for the GGMRF prior distribution, the maximum likelihood (ML) estimate of σ can be shown to have a very simple form [18]:

$$\hat{\sigma}^q = \frac{1}{N} \sum_{\{i,j\}} a_{i,j} \rho(x_i - x_j),$$

where N is the number of pixels in the image.

In X-ray and γ-ray transmission tomography, projection domain noise properties depend on the system's input dosage, whose attenuation provides measurements of integral density. In many systems, the data is stored without preservation of the dosage parameter y_T, whose balance with σ defines the MAP solution. The ML estimate of y_T has also been shown to have a tractable form, and can be computed in parallel to that of σ[19].

The closed form expressions for $\hat{\sigma}$ and $\hat{y}_T$ cannot be directly applied to the tomography problem because the cross-section x is unknown. This type of problem is generally referred to as a missing data problem, and can be solved using the expectation maximization (EM) algorithm in a different setting from its application to MAP reconstruction. For our problem, the EM algorithm results in an iterative procedure in which the estimate of the unknown

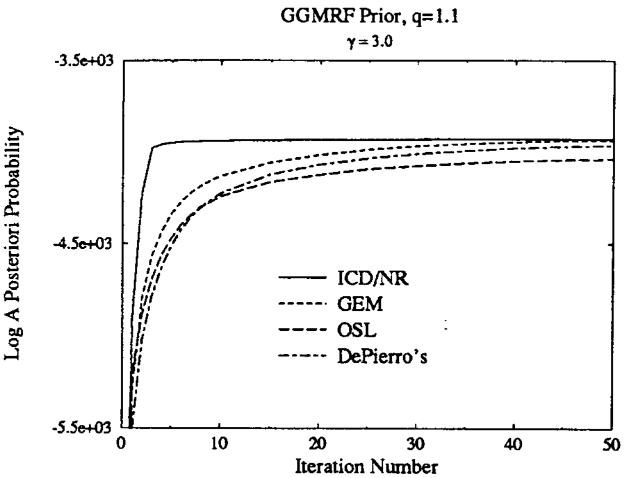

Figure 2: Convergence of MAP estimates with a generalized Gaussian prior model with $q = 1.1$.

parameters $\theta^k = [\sigma^k, y_T^k]^t$ from the observed data $Y = y$ is updated with each iteration k using the formula

$$\theta^{k+1} = \arg\max_{\theta} E\left[\log[g(y, X|\theta)]|Y = y, \theta_k\right] \quad (1)$$

The required conditional expectation can be computed by generating samples from the posterior distribution of the cross-section x given the projection data and θ^k.

5. SPEEDUP TECHNIQUES FOR PARAMETER ESTIMATION

The EM algorithm update of (1) requires the generation of sample reconstructions, X, given the data, y. The Metropolis algorithm is a commonly used simulation method for generating these samples. However, this algorithm tends to suffer from slow convergence to the equilibrium distribution because the transition probability distribution is required to be symmetric.

Hastings [20] and Peskun [21] have 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 $\pi(x)$ be the desired sampling distribution and let $q(x, x')$ be an arbitrary transition probability for generating a new state x' from the current state x. Then the probability of accepting the new state is given by

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

We note that the Gibbs sampler is a special case of this general formulation where the new state for pixel i is generated using conditional distribution, under $\pi(x)$, of x_i given the values of all other pixels. In this case, $\alpha(x, x') = 1$.

Green and Han [22] have argued that convergence is fastest if the transition probability $q(x, x')$ is chosen to be close to that of the Gibbs sampler. This can be done by approximating each pixel's marginal distribution by a Gaussian distribution

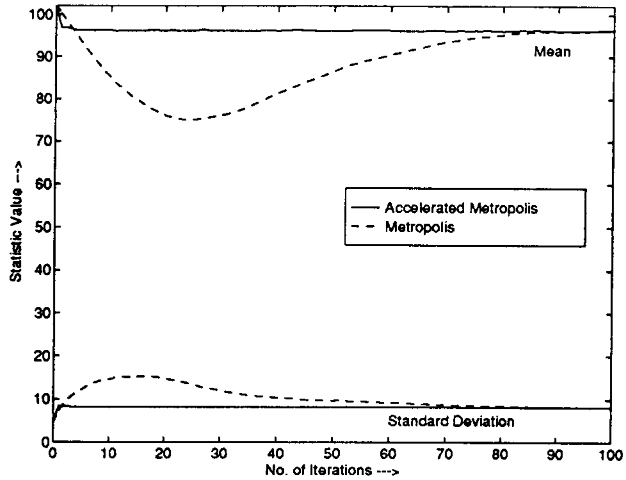

Figure 3: Convergence plot for the mean and standard deviation calculated from the samples X generated from the posterior distribution for the CBP image. A GGMRF with $q = 1.1$ and the positivity constraint was used as the prior distribution of the image.

In the context of the emission tomography, a Gaussian transition probability is not always effective because of the positivity constraint. Therefore, we first compute, μ, the MAP update at a pixel. If the MAP update is positive, then we use a Gaussian transition probability with mean μ and appropriate variance. If the MAP update is negative, then we use a strictly positive exponential distribution with appropriate mean.

Fig. 3 shows the convergence of the image mean and variance calculated from samples of X generated from the posterior distribution. The dotted line shows the conventional Metropolis algorithm while the solid line shows the accelerated Metropolis algorithm outlined above. We see that the new method converges in 2 iterations, while the conventional Metropolis algorithm takes about 70 iterations to converge.

6. CONCLUSION

This collection of techniques for modeling and computation is intended to provide, from at least the signal processing point of view, a relatively complete set of tools for practical applications of statistical tomography. While enjoying the advantages of Bayesian techniques, they require a minimum of operator intervention in parameter choice. They are also flexible enough to easily incorporate a variety of physical models and constraints. The efficiency of the techniques presented offers a step toward practical applications in which computational expense is critical. We hope that further improvements in both algorithms and hardware, and the match between them will bring Bayesian techniques into broader clinical and industrial practice.

7. REFERENCES

[1] S. Geman and D. McClure, "Bayesian Image Analysis:

An Application to Single Photon Emission Tomography," in *Proc. Statist. Comput. Sect. Amer. Stat. Assoc.*, Washington, DC, pp. 12-18, 1985.

[2] H. Hart and Z. Liang, "Bayesian Image Processing in Two Dimensions," *IEEE Trans. Med. Im.*, vol. MI-6, no. 3, pp. 201-208, Sept. 1987.

[3] E. Levitan and G.T. Herman, "A Maximum *A Posteriori* Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography," *IEEE Trans. Med. Im.*, vol. MI-6, no. 3, Sept. 1987.

[4] J. M. Ollinger, "Maximum-Likelihood Reconstruction of Transmission Images in Emission Computed Tomography via the EM Algorithm," *IEEE Trans. Med. Im.*, vol. 13, no. 1, pp. 89-101, March 1994.

[5] P. J. Green "Bayesian Reconstructions from Emission Tomography Data Using a Modified EM Algorithm," *IEEE Trans. Med. Im.*, vol. 9, no. 1, pp. 84-93, March 1990.

[6] K. Lange, "Convergence of EM Image Reconstruction Algorithms with Gibbs Priors," *IEEE Trans. Med. Im.*, vol. 9, no. 4, pp. 439-446, Dec. 1990.

[7] C. Bouman and K. Sauer, "A Generalized Gaussian Image Model for Edge-Preserving MAP Estimation," *IEEE Trans. Image Proc.*, pp. 296-310, vol. 2, no. 3, July 1993.

[8] K. Lange and R. Carson,, "EM Reconstruction Algorithms for Emission and Transmission Tomography," *Journal of Computer Assisted Tomography*, vol. 8, no. 2, pp. 306-316, April 1984.

[9] C. Bouman and K. Sauer, "A Unified Approach to Statistical Tomography Using Coordinated Decent Optimization," *submitted to IEEE Trans. Im. Proc.*

[10] L. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," *IEEE Trans. Med. Im.*, vol. MI-1, no. 2, Oct. 1982.

[11] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Stat. Soc. B*, vol. 39, pp. 1-38, 1977.

[12] L. Kaufman, "Implementing and Accelerating the EM Algorithm for Positron Emission Tomography," *IEEE Trans. Med. Im.*, vol. MI-6, no. 1, pp. 37-51, March 1987.

[13] K. Sauer and C. Bouman, "A Local Update Strategy for Iterative Reconstruction from Projections," *IEEE Trans. Sig. Proc.*, vol. 41, no. 2, pp. 533-548, Feb. 1993.

[14] T. Hebert and R. Leahy, "A Generalized EM Algorithm for 3-D Bayesian Reconstruction from Poisson data Using Gibbs Priors," *IEEE Trans. Med. Im.*, vol. 8, no. 2, pp. 194-202, June 1989.

[15] A.R. De Pierro, "A Modified Expectation Maximization Algorithm for Penalized Likelihood Estimation in Emission Tomography," to appear in *IEEE Trans. on Med. Im.*, Dec. 1994.

[16] J.A. Fessler and A.O. Hero, "Complete-Data Spaces and Generalized EM Algorithms," *Proc. ICASSP 1993*, vol. 5, pp. 1-4, April 27-30, 1993, Minneapolis MN.

[17] J.A. Fessler and A.O. Hero, "Space-Alternating Generalized Expectation-Maximization Algorithm," *IEEE Trans. Sig. Proc.*, vol. 42, no. 10, pp. 2664-2677, Oct. 1994.

[18] C. Bouman and K. Sauer, "Maximum Likelihood Scale Estimation for a Class of Markov Random Fields," *Proc. IEEE Int. Conf. on Acoust. Speech and Signal Proc.* pp. 537-540, Adelaide, South Australia, April 19-22 1994.

[19] K. Sauer and C. Bouman, "Maximum Likelihood Dosage Estimation for Bayesian Transmission Tomography," *Proc. IEEE Int. Conf. on Im. Proc.* pp. 844-848, Austin, TX, Nov. 13-16 1994.

[20] K. Hastings, "Monte Carlo Simulation Methods Using Markov Chains and Their Applications," *Biometrika*, vol. 57, pp. 97-109, 1970.

[21] P. Peskun, "Optimum Monte-Carlo Sampling Using Markov Chains," *Biometrika*, vol. 60, pp. 607-612, 1973.

[22] P. Green, X. Han, "Metropolis Methods, Gaussian Proposals and Antithetic Variables," pp. 142-164 in *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis*, eds. P. Barone, A. Frigessi, and M Piccioni, Springer-Verlag, Berlin, 1990.