

UNCERTAINTY ESTIMATION IN RECONSTRUCTED DEFORMABLE MODELS

K.M. HANSON, G.S. CUNNINGHAM, AND R.J. MCKEE*
Los Alamos National Laboratory, MS P940
Los Alamos, New Mexico 87545 USA

Abstract. One of the hallmarks of the Bayesian approach to modeling is the posterior probability, which summarizes all uncertainties regarding the analysis. Using a Markov Chain Monte Carlo (MCMC) technique, it is possible to generate a sequence of objects that represent random samples drawn from the posterior distribution. We demonstrate this technique for a reconstruction of a two-dimensional object from two orthogonal noisy projections. The reconstructed object is modeled in terms of a deformable geometrically-defined boundary with a constant interior density yielding a nonlinear reconstruction problem. We show how an MCMC sequence can be used to estimate uncertainties in the location of the edge of the reconstructed object.

Key words: Uncertainty estimation, Bayesian estimation, Markov Chain Monte Carlo, deformable geometric model, tomographic reconstruction

1. Introduction

A number of years ago one of the authors (KMH) introduced the use of geometrical models in tomographic reconstruction within a Bayesian framework [1, 2]. Such models, also called deformable models or snakes, can greatly improve one's ability to reconstruct an object from projections taken from a very small number of angles, particularly when the object is known to have constant density. Indeed, it was shown that is possible to obtain a good reconstruction of a simulated lumen (cross section of a blood vessel) from just two orthogonal views that might be obtained from two digitally-subtracted angiograms, for example. Those familiar with tomographic reconstruction problems initially find that result surprising. The astounding success of that demonstration begs the question, how much can one rely on a reconstruction based on such limited data? In this paper we directly address this critical question.

The general method we employ here to estimate uncertainty in the reconstructed object is to generate a sequence of random samples of the posterior probability distribution using the Markov Chain Monte Carlo (MCMC) technique. By fully mimicking the posterior, this sequence of samples can be used to assess the posterior in various ways. We demonstrate posterior characterization for reconstructions of an object from projections in two directions under the assumption

Supported by the United States Department of Energy under contract W-7405-ENG-36.
Email: kmh@lanl.gov, cunning@lanl.gov, mckee@lanl.gov
WWW: <http://home.lanl.gov/kmh>

of a known, constant interior density. In the analysis, the boundary of the reconstructed objects are subject to a prior that promotes smoothness. We show how the MCMC sequence can be used to estimate uncertainties in the location of an edge of the reconstructed object.

Before proceeding, we mention an alternative method of probing the posterior to estimate the uncertainty in specific aspects of an estimated model, which we introduced at the MaxEnt Workshop in 1993 [3, 4]. In that approach, which we often refer to as the “hard truth” concept, one observes the displacement of the most probable model as one applies a specified “force” to the model. We have shown that the resulting displacement in the model parameters is quantitatively related to the covariance matrix times the force vector when the posterior can be approximated by a multivariate Gaussian distribution.

2. Bayesian analysis

Bayesian analysis provides the ultimate means of analysis of uncertainties in the interpretation of data in terms of models. Every aspect of modeling is assigned a probability that indicates our degree of certainty in its value. At the lowest level of analysis, the estimation of the values of parameters for a specified model, a probability density function (PDF) is associated with each continuous parameter. Loosely speaking, the range of a probability distribution indicates the possible range of its associated parameter. The benefit of Bayesian analysis over traditional methods of uncertainty, or error, analysis is that it permits the use of arbitrary probability distributions, not just Gaussian distributions, and of arbitrary measures of uncertainty, not just rms deviation (or variance). Bayesian analysis also brings to light the use of prior knowledge, which all kinds of analysis incorporate, but do not advertise. Furthermore, it extends analysis to higher levels of interpretation, e.g., the choice of hyperparameters that control the strength of the priors used and the selection of appropriate models [5]. We refer the reader to any of the several excellent books that have recently appeared [6, 7, 8], which treat Bayesian analysis from a practical point of view. Another useful source of background information is the collection of proceedings from the series of workshops on Maximum Entropy and Bayesian Methods, which have been published under that title, mostly by Kluwer Academic Press.

In Bayesian analysis, the state of knowledge about the parameters \mathbf{x} associated with a model that describes the physical object being studied is summarized by the posterior, which is the probability density function $p(\mathbf{x}|\mathbf{d})$ of the parameters given the observed data \mathbf{d} . Bayes law gives the posterior as

$$p(\mathbf{x}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{x})p(\mathbf{x}) . \quad (1)$$

The probability $p(\mathbf{d}|\mathbf{x})$, called the likelihood, comes from a comparison of the actual data to the data predicted on the basis of the model of the object. Under the assumption that the data are degraded by uncorrelated and additive Gaussian noise, it is appropriate to use the exponential of $-\frac{1}{2}\chi^2$ for the likelihood. As usual, χ^2 is the mean squared difference between the actual and the predicted

measurements divided by the expected variance in the measurements. The predicted data are generated using a model for how the measurements are related to the object, which we call the measurement model. The prior $p(\mathbf{x})$ expresses what is known about the object, exclusive of the present measurements, and may represent knowledge acquired from previous measurements, specific information regarding the object itself, or simply general knowledge about the parameters, e.g., that they are nonnegative.

2.1. DEFORMABLE MODELS

We will model the objects to be reconstructed in terms of their boundary and their interior density, which is taken to be constant. The boundary is approximated in discrete terms as a finely-divided polygon. The length of the edges of the polygon can be made short enough to adequately describe a curve at any degree of resolution desired. The use of a polygon actually imposes a desirable constraint on the result of an analysis, namely that the object's boundary is closed.

A smoothness constraint on the boundary is achieved by placing a prior on the curvature of the boundary. The minus-log-prior is taken to be proportional to $\int \kappa^2(s) ds$, where $\kappa(s)$ is the curvature of the curve as a function of the position along the curve s . This prior serves to keep the curve smooth because large curvature is penalized. This form for a prior has a physical analog in the formula that describes the potential energy created by bending a stiff rod. We note that since the integral has the dimensions of reciprocal length, it depends on the scale of the curve. To achieve a prior that is related to the shape of the curve, not its size, as suggested in Ref. [9], we suggest that the integral should be multiplied by the total arclength of the boundary to form a dimensionless quantity.

For our discrete polygon model, we replace the integral by a sum of contributions associated with each vertex in combination with half of each neighboring edge of the polygon. To approximate the minus-log-prior for the continuous curve, we use the expression

$$\alpha \frac{1}{2\pi} \left[\sum_j (L_j^- + L_j^+) \right] \left[\sum_j w_j \frac{\tan^2(\theta_j/2)}{L_j^- L_j^+} (L_j^- + L_j^+) \right], \quad (2)$$

where the sums are over the vertices of the polygon, $\pi - \theta_j$ is the internal angle at the j th vertex, L_j^- and L_j^+ are the half-lengths of the previous and next edges of the polygon, and w_j is the weight for the j th vertex. The hyperparameter α (called a hyperparameter rather than a parameter because it controls a general aspect of the model) determines the strength of the prior relative to the likelihood. Further details can be found in [10].

In addition to the smoothness constraint, we have found it useful to control the lengths of the sides of the polygon to avoid any side from either getting too small or much bigger than the rest. This control is accomplished in this study by adding to the above minus-log-prior the following expression

$$\beta \sum_j (L_j^+ - \bar{L}^+)^2, \quad (3)$$

where $\overline{L^+} = \frac{1}{n} \sum_j L_j^+$ is the average half-length of the n sides of the polygon. The choice for the hyperparameter β is somewhat arbitrary. It should not effect the outcome of the modeling effort when chosen over a rather large range of values because this prior only provides control of the polygon representation. The value $\beta = 100$ seems to work well in the present circumstance.

The present study is carried out using the Bayes Inference Engine (BIE). We developed the BIE to allow one to easily develop complex models for both the objects under study and the measurement process. Various aspects of the BIE are described elsewhere [4, 11, 12, 13, 14, 15]. The MCMC technique is a perfect match to the computational approach to Bayesian inference that is the foundation of the BIE.

3. Markov Chain Monte Carlo

The Markov Chain Monte Carlo (MCMC) technique provides a means to sample an arbitrary probability density function. A valuable asset of MCMC is that there are generally no restrictions placed on the PDF and no functional form for the PDF is required. In its minimalist form, MCMC only requires that one be able to calculate $\varphi = -\log(\text{posterior})$, although sometimes the gradient of φ is used.

A Markov chain is a sequence of states in which the probability of each state depends only on the previous state. In MCMC the objective is to generate a sequence of sets of parameters that mimic a PDF, let's call it $q(\mathbf{x})$, where \mathbf{x} is a vector of parameters in the relevant parameter space. More precisely it is desired that the MCMC sequence be in statistical equilibrium with the target PDF $q(\mathbf{x})$, which is achieved when the MCMC sequence is marked by the condition of detailed balance:

$$q(\mathbf{x}) T(\mathbf{x} \rightarrow \mathbf{x}') = q(\mathbf{x}') T(\mathbf{x}' \rightarrow \mathbf{x}) \quad , \quad (4)$$

where $T(\mathbf{x} \rightarrow \mathbf{x}')$ is the transition probability for stepping from \mathbf{x} to \mathbf{x}' . This equation essentially requires that in a very long sequence the number of steps from \mathbf{x} to \mathbf{x}' is identical to the number from \mathbf{x}' to \mathbf{x} . For more information about MCMC, the reader is referred to the recent paper by Besag, et al. [16] or the excellent book by Gilks, et al. [17], which summarizes MCMC.

The MCMC technique [16] makes it feasible to perform some of the difficult technical steps required by probability theory (normalization of PDFs, marginalization, computation of expectation integrals, model selection) in a computer. There is growing optimism among Bayesian advocates that faster computers, together improved computational techniques, will promote the routine use of Bayesian analysis to address the complex models that are demanded in many application areas.

3.1. METROPOLIS ALGORITHM

One of the simplest algorithms used in MCMC calculations is due to Metropolis et al. [18]. This algorithm ensures detailed balance (4) for each step in the sequence. One starts at an arbitrary point in the vector space to be sampled, \mathbf{x}_0 . The general recursion at any point in the sequence \mathbf{x}_k is as follows:

- (1) Pick a new trial sample $\mathbf{x}^* = \mathbf{x}_k + \Delta\mathbf{x}$,
where $\Delta\mathbf{x}$ is randomly chosen from a symmetric probability distribution
- (2) Calculate the ratio $r = q(\mathbf{x}^*)/q(\mathbf{x}_k)$
- (3) Accept the trial sample, i.e. set $\mathbf{x}_{k+1} = \mathbf{x}^*$,
if $r \geq 1$,
or with probability r , if $r < 1$,
otherwise, repeat last point, i.e. set $\mathbf{x}_{k+1} = \mathbf{x}_k$.

Obviously, this algorithm is very simple. The remarkable thing is that it works! However, the computational efficiency of the Metropolis algorithm may be poor when simple trial distributions are used.

3.2. COMPARISON WITH SIMULATION APPROACH

An alternative method that is often used to characterize and test a reconstruction algorithm is based on simulation. The algorithm is used to reconstruct from several sets of simulated data in which new realizations of the noise in the measurements are used in each case. This approach provides insight into the propagation of noise in a reconstruction algorithm. However, it does not fully explore the uncertainty of the reconstruction.

Consider a reconstruction problem in which the measurements are linearly related to the parameters to be estimated; the vector of measurements can be written as the product of a measurement matrix times the parameter vector. In ill-determined reconstruction problems, the measurement matrix possesses a null space, which means that there exist combinations of parameters that do not contribute to the measurements. One way to look at the benefit that one wishes to gain from the use of prior information is that it should fill in the null-space components of the parameter space with something meaningful [19, 20].

It is easy to see that the simulation approach outlined above does not sample the uncertainties in the null space. However, the MCMC approach does so because it randomly samples the parameter space. As such, it can also incorporate uncertainties that arise from the priors.

4. Example

4.1. PROBLEM STATEMENT

We demonstrate the versatility of the Metropolis technique with an example of tomographic reconstruction from just two views. This problem is known generally to be an extraordinarily difficult inverse problem. Its solution is made feasible by employing the prior information that the object being reconstructed has constant density and consists of a fairly simple shape with smooth boundaries. Figure 1 displays the object concocted for this example. It is fashioned to be representative of a lumen, the cross section of an artery, possessing a sizable occlusion. Incidentally, for better visualization this image and all the images shown here are blown up to display the central 80×80 -pixel region instead of their full width of 128×128 pixels. To give the scale of the images, the width and height of the object are roughly 64 pixels.



Figure 1: The original object used in the example.

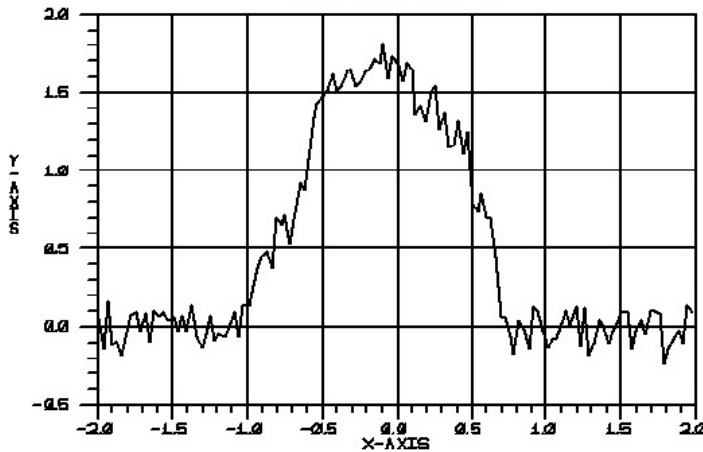


Figure 2: The vertical projection of the original object, with noise added, comprises one of two views used to reconstruct the object.

Two orthogonal views of the object shown in Fig. 1 are generated, one consisting of the vertical projection and the other of the horizontal. Each projection has 128 samples over a distance that is about twice the width of object. Gaussian noise is added to these projections with an rms deviation of 5% of the peak projection amplitude. The data for the vertical projection are shown in Fig. 2.

4.2. MAP RECONSTRUCTION

For reconstruction the object is modeled in terms of a finely-divided polygon filled with a constant density, which we assume is known beforehand. The polygon has 50 sides (and vertices) to fairly approximate a continuous curve. The parameters in the model consist of the x and y values of the 50 vertices. The prior on

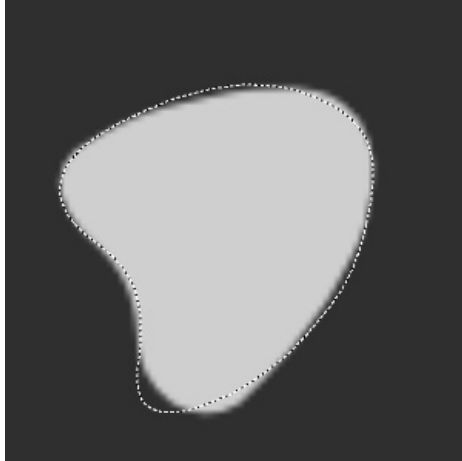


Figure 3: The MAP reconstruction shown as a grayscale image with the boundary of the original object superimposed.

the curvature of the boundary has already been stated. However, the strength of the prior, i.e. α in Eq. (2), must be specified. Ideally, the hyperparameter α that is consistent with the data would be calculated utilizing the next higher level of Bayesian inference [5]. As we are not yet equipped to do that in the BIE, we tried several values for α and selected what seemed to be an appropriate value $\alpha = 1.0$.

The MAP reconstruction is obtained by using the BIE to find the minimum in the minus-log posterior with respect to the 100 variables in the polygon model. The BIE accomplishes this in an efficient manner through the use of the Adjoint Differentiation In Code Technique (ADICT) [21] to calculate the gradient of φ . The reconstructed object compares quite well with the original, as shown in Fig. 3. The maximum discrepancy in the position of the two boundaries is about 3.3 pixels, which occurs in the lower lobe. Over the vast majority of the boundary, the reconstruction lies at most one pixel away from the original.

4.3. MCMC RESULTS

The MCMC algorithm described above was used to generate samples from the posterior of this reconstruction problem. In all, 150,000 trial steps were calculated for a total computation time of about 16 hours on a DECstation 250 with a DEC alpha processor running at 266 MHz. For each MCMC trial step, the increments in the x and y positions of each of the vertices were independently chosen from a Gaussian distribution with an rms step size of 0.06 pixels. 42049 steps were accepted, for an acceptance rate of about 28%. Three widely-separated samples from the full MCMC sequence are shown in Fig. 4. While it is not possible to deduce any quantitative behavior from these three samples, they provide some indication of the amount of variation in the shapes that occupy the posterior. The amount of waviness observed in the boundary is moderate, as can be observed in Fig. 4. The excess amount of waviness compared to the original object might

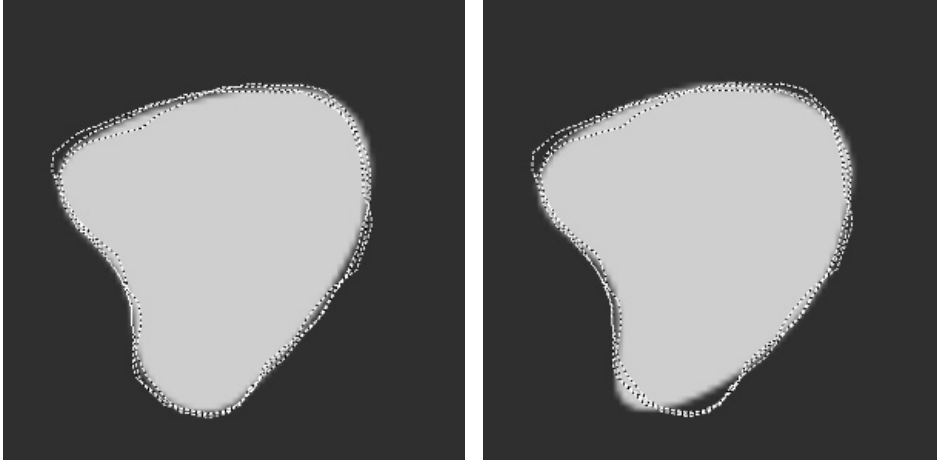


Figure 4: Three representative samples from the posterior shown as curves on top of the grayscale images of the MAP reconstruction (left) and the original object (right).

convince one that $\alpha = 1$ is a safe choice.

During the generation of the MCMC sequence, the configuration of the boundary is saved to disc at every fiftieth step in the sequence. After the full sequence is generated, the saved configurations can be played back as a “video loop”. Visual observation of the abridged sequence indicates that there is a strong correlation in the configurations over a few frames, i.e. more than a hundred steps in the sequence. This correlation might be expected because of the very small step that each vertex takes in each iteration of the Metropolis algorithm. We also observe from the video loop that it takes several hundred steps in the sequence for the boundary to move from far to one side of its mean position to far to the other, a distance of a few pixels. Roughly speaking, one might expect that it would take on the order of $[2/(0.06\sqrt{2})]^2 \approx 500$ random steps of rms radial distance $0.06\sqrt{2}$ pixels to move a total distance of two pixels.

A quantitative estimate of posterior characteristics is obtained by averaging over the MCMC sequence. Such an average of the grayscale image of the object is shown in Fig. 5. Of course, it does not make sense to average the positions of the vertices, because there is nothing to keep the polygon from slipping around the boundary of the object, which has no bearing on the actual object shape. The average MCMC image in Fig. 5 represents the posterior mean image. The amount of blur in the edge of the object indicates the variability in the position of that edge allowed by the posterior, i.e. the uncertainty in edge location. From the measured distance between the 10% and 90% points of the blurred edge of this posterior average, we deduce that the rms uncertainty in edge location varies from about 0.5 pixels to about 1.0 pixels at various positions around the periphery. The smallest rms deviations occur at the limiting edges on the top, bottom, and right sides of the object. These positions are effectively measured by the tangential rays

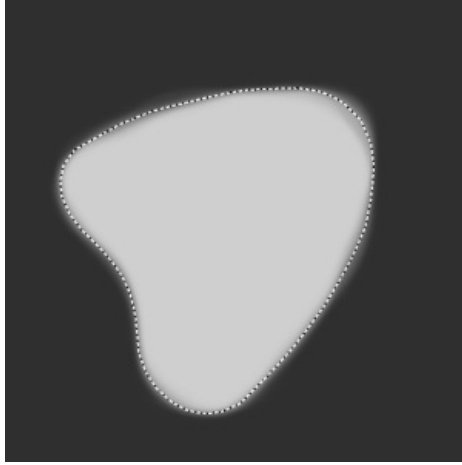


Figure 5: The average of the grayscale images for all the MCMC sample from the posterior with the contour for the MAP reconstruction shown as a dotted line.

of the two projections. The horizontal position of the leftmost edge is not quite as well determined.

One of the largest rms deviations is found to be along the top arc, just left of the middle. This large uncertainty is caused by the ability of the top boundary of the object to easily move left and right and still match the data fairly well. This movement is feasible because it can be matched by small distortions in other parts of the boundary, an effect that can be observed in the video loop. One must be aware that any conclusion drawn from a finite MCMC sequence is subject to uncertainty caused by the statistics of the finite sampling of the posterior. One of the important issues regarding MCMC is how to estimate the accuracy of an uncertainty estimation.

Figure 5 also shows that the MAP reconstruction (the model that maximizes the posterior) appears to be consistent with the contour at half the amplitude of the posterior mean image. This result suggests that the posterior probability distribution is symmetric about its maximum. From the shape of the edge profile of the posterior average, we tentatively conclude that the posterior may be a multivariate Gaussian distribution, despite the nonlinear relation between the vertex parameters and the measurements.

An interesting and important feature of the MCMC technique is that any feature that one wishes to characterize, e.g., the average edge position and its uncertainty in the above example, is not conditional on particular values of other parameters in the model. In terms of probability theory, MCMC provides marginalized results, which means that the dependence on the uncertainties in “nuisance variables” is integrated out. In the context of the above discussion, the uncertainty in the edge position deduced for any particular location of the boundary is independent of the uncertainties in edge positions for the rest of the boundary.

5. Discussion

We have presented the results of a study of MCMC applied to a tomographic reconstruction problem. Although there are an abundance of issues that remain to be addressed, these preliminary results demonstrate the enormous promise of applying MCMC to Bayesian inference. This technique can obviously be used to estimate uncertainties in reconstructed (estimated) models. Moreover, it can be used to estimate the uncertainties in quantities derived from reconstructed models, e.g., in the field of medical imaging, ejection fraction of the heart, or the activity in a specified region of a reconstructed emission image.

On the negative side, many important issues must be addressed before MCMC can be truly useful, including (a) How does one measure the accuracy of an estimate obtained from MCMC? (b) How does one assure adequate coverage of the posterior bubble? (c) How can a minimum number of samples be chosen to represent the posterior in an uncorrelated way? Our simple example leads us to believe the efficiency of the MCMC can be very poor, particularly when the Metropolis algorithm is used with an isotropic trial step distribution. Minimizing the CPU time required to achieve a specified accuracy is of critical importance. One of the approaches we wish to explore soon is to make use of the gradient of the minus-log-posterior, a capability built into the BIE, to better select the MCMC trial steps. The importance of the above topics is reflected by the fact that most of them are currently under active investigation throughout the statistics community. We look forward to innovative solutions coming out of this worldwide research effort.

One of the limitations of the present study is that α in (2) has been fixed. Therefore, the uncertainty in α is not included in the analysis presented above. A more thorough approach would be to consider α as a parameter that should be determined from the data. Then, by including α in the list of variables to be explored by the MCMC process, the uncertainty in α would be included in the overall uncertainty analysis.

Acknowledgments

John Skilling pointed out an inconsistency in the way we define the prior on curvature, but throw MCMC samples in terms of the vertex positions. While we believe this effect to be minor, we will attempt to address this technical issue later. Many other people have lent a hand in helping us understand MCMC and its potential usefulness, including Julian Besag, James Gubernatis, Richard Silver, and Roger Bilisoly.

References

1. K. M. Hanson, "Reconstruction based on flexible prior models," in *Image Processing*, M. H. Loew, ed., *Proc. SPIE*, **1652**, pp. 183–191, 1992.
2. K. M. Hanson, "Bayesian reconstruction based on flexible prior models," *J. Opt. Soc. Amer. A*, **10**, pp. 997–1004, 1993.
3. K. M. Hanson and G. S. Cunningham, "The hard truth," in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., pp. 157–164, Kluwer Academic, Dordrecht, 1996.

4. K. M. Hanson and G. S. Cunningham, "Exploring the reliability of Bayesian reconstructions," in *Image Processing*, M. H. Loew, ed., *Proc. SPIE*, **2434**, pp. 416–423, 1995.
5. D. J. C. MacKay, "Bayesian interpolation," *Neural Computation*, **4**, pp. 415–447, 1992.
6. D. S. Sivia, *Data Analysis: A Bayesian Tutorial*, Clarendon, Oxford Univ. Press, New York, 1996.
7. J. J. K. Ó Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*, Springer, New York, 1996.
8. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, London, 1995.
9. S. Lobregt and M. A. Viergever, "A discrete dynamic contour model," *IEEE Trans. Med. Imaging*, **MI-14**, pp. 12–24, 1995.
10. K. M. Hanson, R. L. Bilisoly, and G. S. Cunningham, "Kinky tomographic reconstruction," in *Image Processing*, M. H. Loew and K. M. Hanson, eds., *Proc. SPIE*, **2710**, pp. 156–166, 1996.
11. K. M. Hanson, G. S. Cunningham, G. R. Jennings, Jr., and D. R. Wolf, "Tomographic reconstruction based on flexible geometric models," in *Proc. IEEE Int. Conf. Image Processing, vol. II*, pp. 145–147, IEEE, 1994.
12. K. M. Hanson and G. S. Cunningham, "The Bayes inference engine," in *Maximum Entropy and Bayesian Methods*, K. M. Hanson and R. N. Silver, eds., Kluwer Academic, Dordrecht, 1995 (to be published).
13. G. S. Cunningham, K. M. Hanson, G. R. Jennings, Jr., and D. R. Wolf, "An object-oriented optimization system," in *Proc. IEEE Int. Conf. Image Processing, vol. III*, pp. 826–830, IEEE, 1994.
14. G. S. Cunningham, K. M. Hanson, G. R. Jennings, Jr., and D. R. Wolf, "An interactive tool for Bayesian inference," in *Review of Progress in Quantitative Nondestructive Evaluation*, D. O. Thompson and D. E. Chimenti, eds., vol. 14A, pp. 747–754, Plenum, New York, 1995.
15. G. S. Cunningham, K. M. Hanson, G. R. Jennings, Jr., and D. R. Wolf, "An object-oriented implementation of a graphical-programming system," in *Image Processing*, M. H. Loew, ed., *Proc. SPIE*, **2167**, pp. 914–923, 1994.
16. J. Besag, P. Green, D. Higdon, and K. Mengersen, "Bayesian computation and stochastic systems," *Stat. Sci.*, **10**, pp. 3–66, 1995.
17. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, 1996.
18. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machine," *J. Chem. Phys.*, **21**, pp. 1087–1091, 1953.
19. K. M. Hanson and G. W. Wecksung, "Bayesian approach to limited-angle reconstruction in computed tomography," *J. Opt. Soc. Amer.*, **73**, pp. 1501–1509, 1983.
20. K. M. Hanson, "Bayesian and related methods in image reconstruction from incomplete data," in *Image Recovery: Theory and Application*, H. Stark, ed., pp. 79–125, Academic, 1987.
21. K. M. Hanson and G. S. Cunningham, "A computational approach to Bayesian inference," in *Computing Science and Statistics*, **27**, M. M. Meyer and J. L. Rosenberger, eds., pp. 202–211, Interface Foundation, Fairfax Station, VA 22039-7460, 1996.