

# Kinky tomographic reconstruction

Kenneth M. Hanson<sup>a</sup>, Roger L. Bilisoly<sup>b</sup>, and Gregory S. Cunningham<sup>a</sup>

<sup>a</sup>Los Alamos National Laboratory, MS P940, Los Alamos, NM 87545

<sup>b</sup>National Center for Atmospheric Research, PO Box 3000, Boulder, CO 80307-3000

## ABSTRACT

We address the issue of how to make decisions about the degree of smoothness demanded of a flexible contour used to model the boundary of a 2D object. We demonstrate the use of a Bayesian approach to set the strength of the smoothness prior for a tomographic reconstruction problem. The Akaike Information Criterion is used to determine whether to allow a kink in the contour.

**Keywords:** active contours, deformable geometry, Akaike Information Criterion, Bayesian analysis, model-order selection

## 1. INTRODUCTION

Deformable geometry, mostly in the form of active contours, is increasingly being used in image analysis. The motivation for using this type of representation is that it guarantees a closed curve, which can define the boundary of an object of interest. Furthermore, the smoothness of the curve can be controlled by using a regularizer. The present work is directed toward understanding how strongly the constraint of smoothness should be enforced. Pushing this issue to the logical limit, we also consider how to decide when a smooth curve should be allowed to develop a kink, that is, an abrupt change in slope.

The context of this paper is computed tomography – reconstruction of a 2D density distribution from sets of line integrals taken through the unknown distribution. We will concentrate on the underdetermined problem of reconstruction from projections measured from a limited number of directions. This kind of tomographic reconstruction problem can be solved in the appropriate circumstances by modeling the object in terms of a warpable geometry.<sup>1-3</sup> One way to describe the object is in terms of a constant interior density enclosed by a flexible boundary. A generic problem arises, however: How does one control the assumed smoothness of the boundary in an optimal way? One would like to determine the smoothness from the data, when they are informative enough. So the issue finally comes down to quantifying the information contained in the data relative to the position and smoothness of a contour, which is an issue relevant to almost every analysis that employs deformable geometry.

In this paper we determine the strength of the smoothness regularization using a Bayesian approach. We follow the methodology described by MacKay,<sup>4</sup> who expanded on the work of Gull.<sup>5</sup> To make decisions about whether to create a kink in the contour, we explore using the Akaike Information Criterion<sup>6</sup> (*AIC*) to make the decision about when to allow the smoothness constraint to be broken. The Akaike approach is based on the balance between the information carried by the data and the number of degrees of freedom used in a model to describe the thing of interest. In a sense the Akaike criterion does not take prior knowledge into account. A fully Bayesian approach to this question would be preferable, of course.

## 2. BAYESIAN METHODOLOGY

The Bayesian approach provides a thorough methodology for building models from measurements. Given a set of measurements, it can be used to answer the fundamental questions: What are the “best” values of the parameters for a particular model? What is the uncertainty in the parameter values implied by uncertainty in the measurements? Which of two models is most appropriate to use? If using prior information, how strongly should one rely on it? The reader is encouraged to learn more about the fundamentals of Bayesian analysis by studying some of the references available.<sup>4,5,7,8</sup>

---

Further author information –

K.M.H.(correspondence): Email: kmh@lanl.gov; WWW: <http://www.lanl.gov/home/kmh/>; Telephone: 505-667-1402; Fax: 505-665-3359

G.S.C.: Email: cunning@lanl.go; Telephone: 505-667-2562

R.L.B.: bilisoly@ucar.edu; Telephone: 303-497-1334

## 2.1. The Basics

In Bayesian analysis, the posterior probability summarizes the full state of knowledge concerning a given situation. Given the data  $\mathbf{d}$ , the posterior probability of parameters  $\mathbf{a}$  for a particular model  $M_l$  is given by Bayes' law

$$p(\mathbf{a}|\mathbf{d}, M_l) = \frac{p(\mathbf{d}|\mathbf{a}, M_l) p(\mathbf{a}|M_l)}{p(\mathbf{d}|M_l)}, \quad (1)$$

where  $p(\mathbf{d}|\mathbf{a}, M_l)$ , the probability of the observed data given  $\mathbf{a}$ , is called the likelihood,  $p(\mathbf{a}|M_l)$  is the prior probability of  $\mathbf{a}$ , and  $p(\mathbf{d}|M_l)$  is called the evidence. The likelihood is specified by the probability distribution of the fluctuations in the measurements about their predicted values. The prior probability  $p(\mathbf{a}|M_l)$  encompasses the full prior information about the relative frequency of occurrence of all possible parameters. Since the evidence can be written as  $p(\mathbf{d}|M_l) = \int p(\mathbf{d}|\mathbf{a}, M_l) p(\mathbf{a}|M_l) d\mathbf{a}$ , it provides the normalization for the posterior. It is usually not necessary to consider the evidence in making inferences about the parameters. However, as it does depend on the model used, the evidence plays a central role in making decisions about the correct model, including the choice of regularization parameters.

It is convenient to deal with the negative logarithm of the posterior:

$$-\log[p(\mathbf{a}|\mathbf{d})] = \varphi(\mathbf{a}) = \Lambda(\mathbf{a}) + \Pi(\mathbf{a}) - E, \quad (2)$$

where  $\Lambda$ ,  $\Pi$ , and  $E$  are the negative logarithms of the likelihood, the prior, and the evidence, respectively.  $\Lambda$  and  $\Pi$  depend on the parameters, while  $E$  does not. We often seek to find the parameter values that maximize the posterior, called the MAP solution. Of course, the MAP estimate is found by minimizing  $\varphi$  with respect to the parameters, yielding the estimated parameter values  $\hat{\mathbf{a}}$ . The condition for the MAP solution is  $\frac{\partial \varphi}{\partial a_i} = 0$  for all parameters  $a_i$ , providing there are no constraints on the parameters themselves.

Under the assumption that the measurements are subject to additive, uncorrelated Gaussian noise, the minus-log-likelihood is half of the familiar chi squared,  $\Lambda(\hat{\mathbf{a}}) = \frac{1}{2}\chi^2 = \frac{1}{2}\sum_i \sigma_i^{-2}(d_i - \hat{d}_i)^2$ , the sum of the squared residuals (the difference between an observed measurements and their values predicted by the estimated parameters  $\hat{\mathbf{a}}$ ) divided by the estimated variance of the noise,  $\sigma^2$ .

It is often useful to draw an analogy between the negative log-probabilities, such as  $\varphi$ , and a physical potential. Such an analogy can be used to give insight into a meaningful form of a prior,<sup>1,2</sup> as we do in the next section. This analogy has also led us to a new way to explore the uncertainties of parameters estimated in a Bayesian analysis.<sup>9,10</sup> From this analogy it is apparent that minus the derivative of  $\varphi$  with respect to a set of parameters represents a force that exerts itself on those parameters. This force points the way towards higher probability, as judged by the data and the prior information. Likewise, the negative derivative of  $\Lambda$  represents the force on the parameters that arises from the data alone. When  $\varphi$  can be approximated by a quadratic, as it often can, the force that it generates is proportional to the displacement of the parameters away from their MAP values, much like a spring in mechanics. With this in mind, the uncertainty of a solution can be seen to be related to the ‘‘stiffness’’ of the model.

We will show how to employ this analogy between minus-log-probabilities and potentials to understand how a model of the reconstructed object needs to change to match the data.

## 2.2. Choosing the Strength of the Prior

The Bayesian approach provides a means to choose models, as summarized by MacKay.<sup>4</sup> For example, it is possible to determine so-called hyperparameters, which include parameters that control the strength of priors. Suppose that the prior is written as  $\Pi(\mathbf{a}) = \alpha C(\mathbf{a})$ , where the function  $C$  is used to characterize the prior on the parameters  $\mathbf{a}$ . If the value of  $\alpha$  is not known beforehand, we need to infer it from the data using Bayes' law:

$$p(\alpha|\mathbf{d}, C, M_l) = \frac{p(\mathbf{d}|\alpha, C, M_l) p(\alpha|C, M_l)}{p(\mathbf{d}|C, M_l)}, \quad (3)$$

where  $p(\mathbf{d}|\alpha, C, M_l)$  is the likelihood for  $\alpha$ ,  $p(\alpha|C, M_l)$  is the prior on  $\alpha$ , and  $p(\mathbf{d}|C, M_l)$  is the evidence for the data given the form for the prior  $C$ ,  $p(\mathbf{d}|C, M_l) = \int p(\mathbf{d}|\alpha, C, M_l) p(\alpha|C, M_l) d\alpha$ . This latter quantity is not important for inferring  $\alpha$ . The likelihood of  $\alpha$  is of course the same as the evidence for the data given  $\alpha$ ,  $C$ , and the object model  $M_l$ , which is the denominator in Eq. (1) but with the  $\alpha$  and  $C$  dependences explicitly shown. We thus see how the

normalizing term of Bayes' law for parameter estimation becomes the essential factor in making inferences about the model, in this case for the prior.

In his derivation for the appropriate value for  $\alpha$ , MacKay assumes that the prior on  $\alpha$  is flat. The MAP estimate for  $\alpha$  then occurs at the maximum in the evidence in Eq. (1). By assuming that the minus-log-posterior for  $\mathbf{a}$ , given by Eq. (2), can be approximated by a quadratic function with a curvature matrix  $\mathbf{A}$ , MacKay shows how to obtain Gull's result for the condition to estimate  $\alpha$ :

$$\alpha C(\hat{\mathbf{a}}) = \gamma(\alpha) , \quad (4)$$

where  $\gamma(\alpha) = k - \alpha \text{Trace}[\mathbf{A}(\alpha)^{-1}]$  is called the number of good parameter measurements, in the case of  $k$  parameters  $\mathbf{a}$ . The quantity  $C(\hat{\mathbf{a}})$  is the prior evaluated at the MAP estimate for a specific value of  $\alpha$ .

The parameter  $\gamma$  is essentially the number of degrees of freedom in the model that are effectively determined by the data rather than by the prior. For very large values of  $\alpha$ ,  $\gamma$  approaches zero, because the prior eventually overwhelms the likelihood. As  $\alpha$  goes to zero,  $\gamma$  approaches  $k$ , because the prior becomes too weak to affect the parameter values. Equation (4) is an implicit equation that must be solved for the most probable value of  $\alpha$ . Skilling<sup>11</sup> has shown how one can estimate the  $\text{Trace}[\mathbf{A}(\alpha)^{-1}]$  in (4) for very large matrices using a Monte Carlo method. We will discuss how to use this expression to choose the strength of the smoothness prior on the deformable boundary of an object.

### 2.3. The Akaike Information Criterion

MacKay also showed how the Bayesian approach can be used to select the appropriate model. However, we shall try to avoid the detailed Bayesian calculation and instead rely on the Akaike Information Criterion<sup>6,12</sup> (*AIC*)

$$AIC = 2\Lambda(\hat{\mathbf{a}}) + 2k , \quad (5)$$

where  $k$  is the number of parameters in the model and  $\hat{\mathbf{a}}$  is the parameter set that minimizes  $\Lambda$ . According to the *AIC*, one accepts a new parameter if the *AIC* drops with the addition of the new parameter. This criterion can be used to determine when the data suggest that the boundary of our reconstructed object should depart drastically from being smooth. The *AIC* is closely related to a Bayesian approach.<sup>13</sup> Since typically  $2\Lambda = \chi^2$ , this criterion is comparable to the rule-of-thumb that one may add a new parameter to the model as long as the minimum in  $\chi^2$  drops by more than one. However, the *AIC* is more conservative than that rule of thumb because it requires that  $\chi^2$  drop by at least two. If the *AIC* is not met, the new parameter is needlessly adding complexity to the model. The *AIC* thus plays the role of Occam's razor, which is the Bayesian way of stating our preference for parsimonious descriptions of reality.

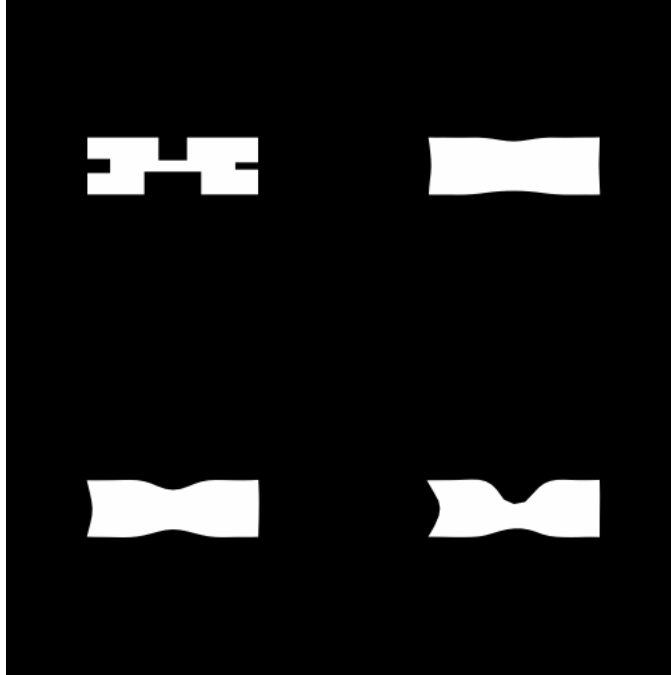
## 3. DEFORMABLE CONTOURS

We will model objects to be reconstructed in terms of their boundary and their interior density, which will often be taken to be constant. A continuous contour can be 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 any degree of resolution desired. The use of a polygon actually imposes a significant desirable constraint on the result of an analysis, namely that the object's boundary is closed.

### 3.1. Continuous Model

We wish to place a prior on contours to prevent them from becoming unnecessarily jagged. To impose such smoothness on a continuous model of a contour, we use what has become a standard choice for which the minus-log-prior is taken to be proportional to  $\int \kappa^2(s) ds$ , where  $\kappa(s)$  is the curvature of the curve. This prior serves to keep the curve smooth because severe 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 units of length in which it is calculated. To overcome this problem, we suggest that the integral be multiplied by the total arclength of the boundary  $S$  to form a dimensionless quantity. This quantity is related to the shape of the curve, not its size, as suggested in Ref. 14. For reasons that will become apparent later, we suggest an additional normalizing factor of  $(2\pi)^{-1}$ .

A noteworthy extension to the above approach to smoothing is to allow the strength of the constraint to vary with position around the boundary. This is simply achieved by weighting the integrand of the curvature integral by a function  $w(s)$ , so that it becomes  $\int w(s) \kappa^2(s) ds$ .



**Figure 1.** A composite of four images. The upper-left section shows the test object used to demonstrate the effect the smoothing prior has on the boundary of a reconstructed object. The objective is to match this shape with an object defined in terms of its flexible boundary. The upper-right shows the reconstructed object obtained for a relatively large value of the regularization parameter,  $\alpha$ . The smoothing constraint is turned off at the corners. The result of increasing  $\alpha$  by successive factors of two is shown in the lower left and the lower right, respectively.

### 3.2. Discrete Model

For our discrete 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. We approximate the curvature by

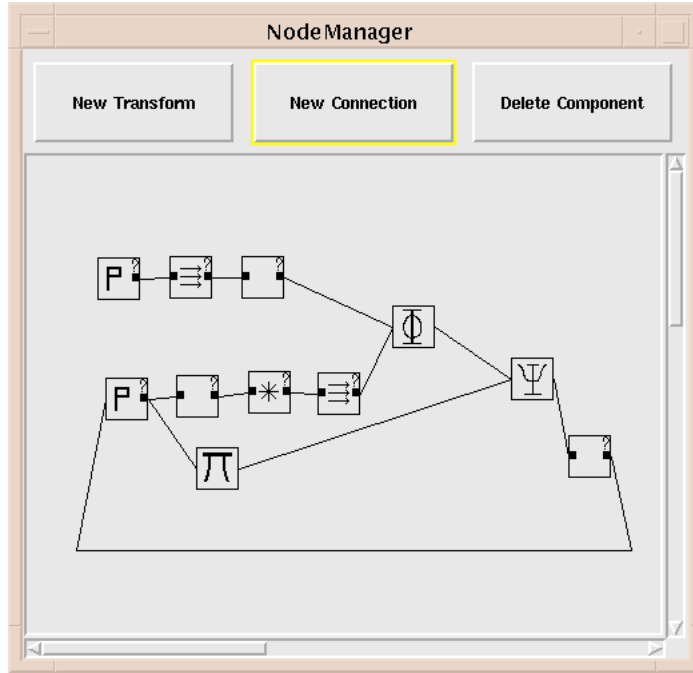
$$\kappa = \frac{\tan(\frac{1}{2}\theta)}{\sqrt{L_1 L_2}}, \quad (6)$$

where  $\theta$  is the angle between the two edges meeting at the vertex and  $L_1$  and  $L_2$  are the half lengths of the two edges. When  $L_1 = L_2$ , this formula agrees with the curvature of a circle that is tangent to the two edges at their midpoint. When  $L_1$  is not the same as  $L_2$ , the  $\kappa$  obtained is the geometric mean of the curvatures of two circles, one for a corner of half-edge length  $L_1$  and the other for  $L_2$ . One desirable consequence of Eq. (6) is that  $\kappa$  goes to infinity when an edge folds back on its neighbor. This type of action mimics what we would like, because it strongly discourages kinky behavior. For the discrete version of  $ds$ , we simply use  $\Delta s = L_1 + L_2$ . Consequently, the minus-log-prior on the curvature becomes

$$C = \frac{1}{2\pi} \left[ \sum_j (L_j^- + L_j^+) \right] \left[ \sum_j w_j \frac{\tan^2(\theta_j/2)}{L_j^- L_j^+} (L_j^- + L_j^+) \right], \quad (7)$$

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.

Figure 1 demonstrates the effect of the above prior on curvature. The objective is to match the shape in the upper-left panel, a rectangle with slots of various widths, by an object defined as unity inside a closed curve and zero outside. The initial shape of the boundary is a rectangle, which is achieved by using a polygon in which each of the long straight sides of the rectangle are comprised of 24 edges and each of the short sides of 8 edges. The



**Figure 2.** The data-flow diagram for the Bayes Inference Engine that is used in the present tomographic reconstruction problem.

weight function  $w_j$  in the constraint sum is set to zero at each of the four corners. In this mock problem, we use for  $\chi^2$  the sum of the squares of the difference between the target shape and the active object. The other panels in the figure show how the sides of the rectangular object are sucked into the slots in the original figure. From the analogy between this example and the simple bending of beams, we expect the constant pressure on the stiff contour would result in the contour reaching a state of constant curvature within each slot, independent of the width of the slot. Furthermore, the curvature of the boundary should be inversely proportional to the strength of the prior, i.e.  $\alpha$ . These conditions are qualitatively met in this example. Deviations from this expected behavior probably occur because of the discrete approximations used for the image in terms of pixels and the contour, i.e. polygon.

#### 4. THE BAYES INFERENCE ENGINE

The Bayes Inference Engine (BIE) is used in this study to create and analyze the data from the simulated tomographic reconstruction problem, which will be discussed in the next section. The BIE is a versatile application, developed by us at the Los Alamos National Laboratory, that allows one to easily develop complex models for both the objects under study and the measurement process. Various aspects of the BIE are described elsewhere.<sup>3,10,15–18</sup>

##### 4.1. Data-flow Diagram

The user interface for the BIE is a graphically programmed data-flow diagram. The data-flow diagram used for this study is shown in Fig. 2. The boxes on the top row, from left to right, represent the original image (to be reconstructed) followed by the transforms to take the projections of that image and to add noise to the projections. The result of this simulation of the measurements is fed into the minus-log-likelihood module (the box with the  $\Phi$  inside it). The second row comprises the model of the object used to analyze the data. The leftmost box represents the finely-divided polygon used to define the boundary of the object. This polygon is converted to a pixelated image, which is  $256 \times 256$  in size, in the next box. In this conversion, it is assumed that the inside of the polygon has a constant density of unity and the density outside is zero. This image is then blurred by convolution with a Gaussian whose width can be adjusted. The reason for this operation will be described momentarily. The last box on that row takes the projections of the blurred image and feeds the projections into the likelihood.

The minus-log-likelihood is calculated here as  $\frac{1}{2}\chi^2$ . The minus-log-prior, which is based on the boundary model, is represented by the box with a  $\Pi$  in it. The value of the minus-log-likelihood is added to the minus-log-prior in the box labeled  $\Psi$  to yield the minus-log-posterior. The minus-log-posterior is minimized by the box on the right, called the Optimizer, by adjusting the parameters in the boundary model. The line at the bottom of the diagram indicates this feedback.

## 4.2. Adjoint Derivatives of $\varphi$

A unique feature of the BIE is its ability to provide derivatives of  $\varphi$  with respect to model parameters at any point in an arbitrary data-flow diagram. This remarkable feat is achieved by combining the technique of adjoint differentiation with object-oriented programming.<sup>15</sup> In adjoint differentiation the derivatives of a scalar quantity at the end of a sequence of transformations are propagated through the data-flow diagram in the reverse direction. This process merely amounts to an implementation of the well-known chain rule of differentiation. Each transformation in a BIE model can display its own adjoint derivative with respect to its input data, as well as its own output data.

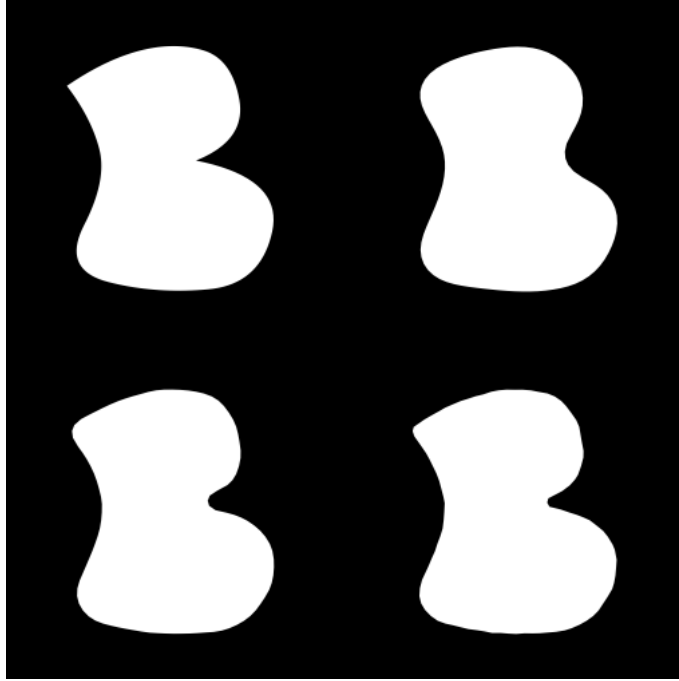
This feature allows us to view the gradient of  $\varphi$  with respect to every aspect of the model. For example, the log-likelihood module (the box with the  $\Phi$  on it) can display the derivative of the log-posterior (in order to be confusing, the box with the  $\Psi$  on it) with respect to its inputs, namely the projection data that are compared in calculating it. This derivative is simply the residuals, the difference between the measurements and their values predicted by the BIE model. Working backward along the second row of the diagram in Fig. 2, the projection transform can show the gradient of  $\varphi$  with respect to its input image, which is the backprojection of the residuals (coming from the adjoint output of the log-likelihood module). Going back one more module the adjoint derivative shown by the convolution transform is the convolution of the backprojected residuals with the same (actually adjoint) blur function as used in the forward calculation. Ultimately, we end up at the beginning of the formulation of the forward model with the specification of the object's boundary. This module can show the derivatives of  $\varphi$  with respect to the vertices of the polygon as scalable vectors emanating from each vertex.<sup>19</sup> This convenient method of visualizing the gradients of  $\varphi$ , coupled with the interpretation of these gradients as forces on the parameters being generated by the data, thorough the log-likelihood, or by the prior, provides a powerful paradigm for understanding the details of the analysis and the interplay of conflicting demands on the model.

The blurring module in the middle of the second row of the data-flow diagram is included to facilitate the multiscale approach to fitting contours, as described by Cunningham et al. in these Proceedings.<sup>19</sup> The essential idea is that the forces acting on the contour to bring it into agreement with the data become more homogeneous and make them approximately linear with distance of the contour from the position demanded by the data. The blurring effectively produces multiscale derivatives. A large amount of blur is used at the beginning of the optimization process to avoid uneven displacements of the curve when it is far from the final shape. The blur is decreased as the final shape is approached.

## 5. RESULTS FOR A TOMOGRAPHIC PROBLEM

We will demonstrate how the Bayesian approach can be used to determine the strength of the curvature prior in a tomographic reconstruction problem. The object that we wish to reconstruct is shown in the upper-left portion of Fig. 3 as a  $256 \times 256$  pixellated image. It is assumed that six parallel projections of the object have been measured at an angular spacing of 30 degrees. At each angle, 256 evenly-spaced line integrals are measured. The measurements are degraded by adding to them uncorrelated Gaussian noise with an rms value of 0.02, which is about 1% of the largest projection value. The knowledge that the object is reasonably compact and has a constant density within its fairly simple boundary is to be exploited to make this problem soluble. The object is modeled in terms of a deformable boundary, within which the density is constant.

The model implemented in the BIE to analyze these data was discussed in the previous section. The minus-log-posterior is given by Eq. (2) with the minus-log-prior given by  $\Pi = \alpha C$ , where  $C$  is the constraint on the curvature from Eq. (7). The polygon model used to represent the boundary of the object has 100 edges and the interior density is assumed to be known. The initial shape of the object is chosen to be a circle, of unit radius, which is the default shape for our prior.



**Figure 3.** The upper-left panel shows the original object to be reconstructed from six slightly noisy projections. The other three panels show the reconstructions obtained for specific values of the parameter  $\alpha$ ; 250 for the upper right, 7.5, lower left, and 0.75, lower right.

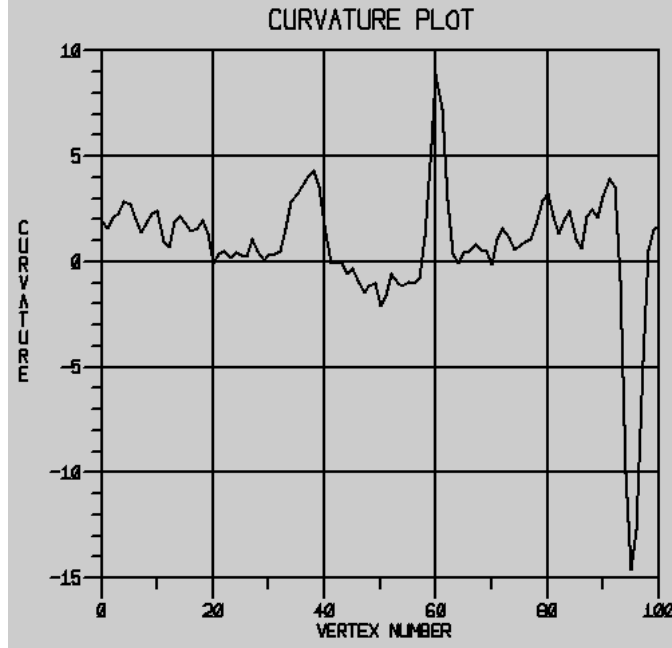
**Table 1.** Summary of reconstructions of the object for several fixed values of  $\alpha$ .

$\alpha$	$\Lambda = \frac{1}{2}\chi^2$	$\Pi = \alpha C$	$C$	$\hat{\gamma}$
250	2950.2	6873.3	27.5	24
25	1150.2	1160.3	46.4	56
7.5	899.0	443.5	59.1	80
2.5	824.3	179.2	71.7	100
0.75	795.8	65.4	87.2	112

### 5.1. Reconstruction Without Kinks

Our first effort is to obtain the best reconstruction that we can employing just the prior on curvature with no kinks allowed. Using the model for the object described in the previous sections, we use the Optimizer of the BIE to find the contours that minimize  $\varphi$  for fixed values of  $\alpha$ . The resulting objects provided by the BIE are shown in Fig. 3 for three different values of  $\alpha$ . The three reconstructions resemble the original object more and more as  $\alpha$  decreases. However, for the smallest value of  $\alpha$ , the boundary starts to get a little ragged, which is just detectable in this figure. The quantitative values obtained for the important quantities are given in Table 1 for five values of  $\alpha$ . We can see from the rise in  $C$  that the boundary indeed develops a larger mean-square curvature as  $\alpha$  decreases.

To fully implement Gull’s Bayesian criterion for determining the best value for  $\alpha$ , we would need to evaluate  $\gamma$ . While this calculation is not impossible, we do not yet have the means to do so. However, given the interpretation of  $\gamma$ , we can roughly estimate  $\gamma$  for a contour by counting the number of degrees of freedom that appear to be dominated by the data for each  $\alpha$ . Figure 4 shows a plot of the curvature of the boundary of the reconstructed object for  $\alpha = 2.5$ . The default curvature for our prior is constant with a value of one, because we start with a circle of unit radius. Therefore, any deviation from a constant value of curvature must be caused by the data. We argue that we can get a reasonable estimate for  $\gamma$  by determining the minimum number of parameters needed to account for



**Figure 4.** A plot of the curvature of the reconstructed boundary versus the vertex index for  $\alpha = 7.5$ . The value of  $\gamma$  is estimated from the number of peaks in this plot.

the structure in the curvature plot. If we consider each bump in the curve as a quadratic function of arclength, then we would need to use a Bézier patch of fourth degree to represent the contour associated with that bump. To avoid discontinuities in the slope of the curvature plot, we would apply  $C^3$  constraints across patch boundaries. Then each bump (patch) would have two free parameters. Therefore, we arrive at an estimate for  $\hat{\gamma}$  of four times the number of distinct local maxima in the curvature plot. For Fig. 4 we count 20 peaks to obtain  $\hat{\gamma} \approx 80$ . The estimated values of  $\hat{\gamma}$  for all the reconstructions are given in Table 1.

From this analysis, Gull’s condition,  $\alpha C = \gamma$  is seen to be achieved at  $\alpha \approx 1.4$ . From the appearance of the results shown in Fig. 3 for  $\alpha = 0.75$ , this result seems acceptable, although a larger value of  $\alpha$  might be preferable because the noise is starting to unduly affect the boundary. Of course, we should keep in mind that our preference for a smoother boundary over one that is beginning to show the effects of noise in the measurements can be incorporated in Bayesian analysis through an appropriate adjustment of the prior on  $\alpha$ . We conclude that Gull’s condition provides a reasonable way to choose the strength of a prior.

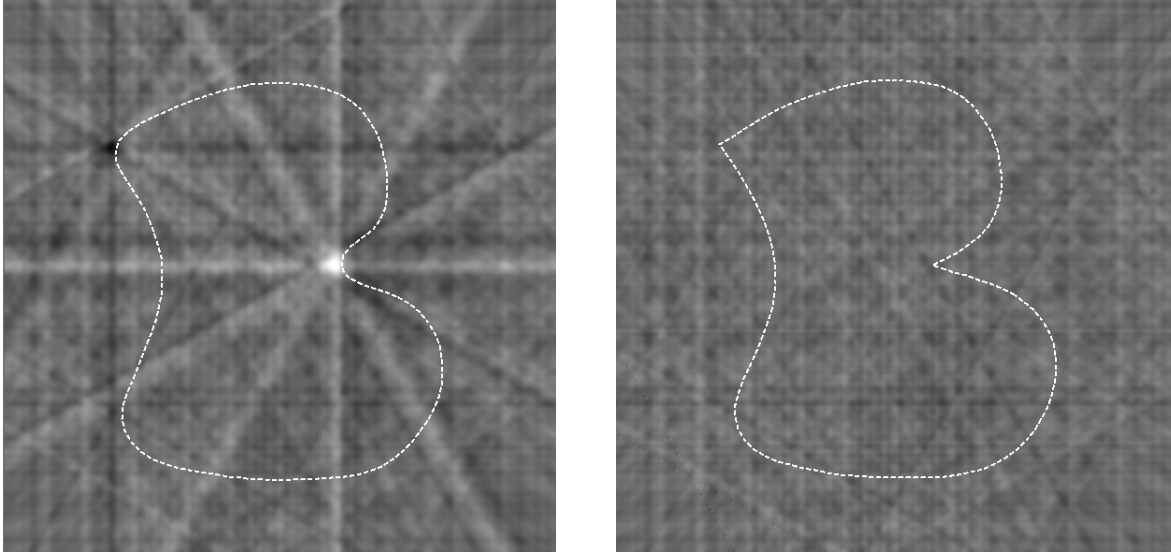
## 5.2. The Case For Kinks

Having performed the analysis outlined above, we should ask whether the data are adequately matched by our final model. The value of  $\chi^2$  for  $\alpha = 0.75$  is 1596, which is about the same as the number of measurements\*, 1536. We might conclude that the data are matched well enough. However, a better approach is to look at the residuals. We can do this in the domain of the problem, rather than in that of the measurements by displaying the adjoint derivative of  $\varphi$  with respect to the reconstructed image densities. That is, we view the backprojected residuals, as shown in Fig. 5 for  $\alpha = 25$ . We see in the left-hand image that there is a distinctly coherent pattern suggesting that the reconstruction is deficient in two places, which are at regions of very high curvature in the boundary. It seems that the data are trying to tell us that even higher curvature is desirable. Indeed, we know that as the stiffness of the boundary is decreased by lowering  $\alpha$ , the curvature in these regions keeps increasing. In our mechanical analogy, the data are trying to “break” the curve by bending it back on itself. Thus we are led to consider the possibility of a kink at these locations.

---

\*In this kind of problem, one would normally expect to achieve a value of  $\chi^2$  that is less than the number of measurements. For the particular data set used in this study, the rms noise value is actually 1.028 instead of unity, so that the sample value of  $\chi^2$  when the true model is used is 1622. The minimum value of  $\chi^2$  is indeed smaller than this value.





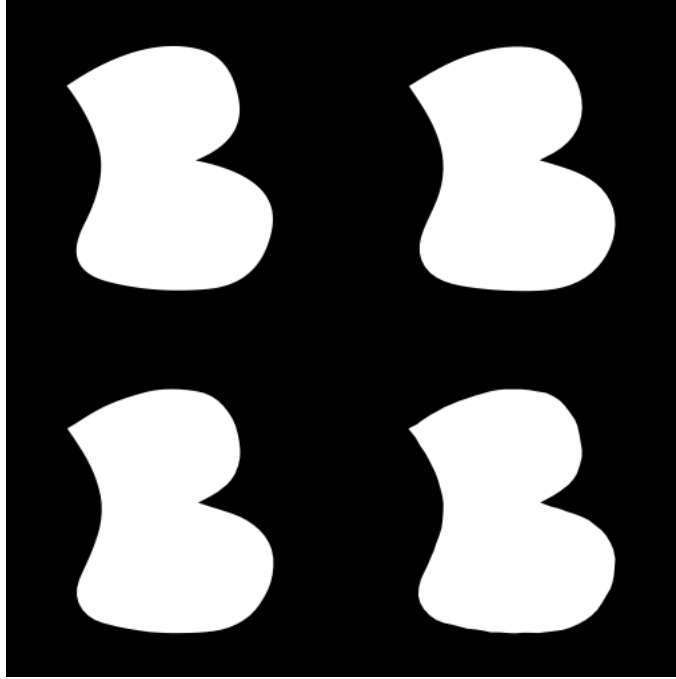
**Figure 5.** The image on the left shows the contour reconstructed for  $\alpha = 25$ , superimposed on the backprojected residuals, which indicate that the data demand a decrease in density at the apex of the indentation in the contour and an increase in density at the upper-left apex. The result of allowing a kink to develop at these two points is shown on the right; the backprojected residuals now do not appear to pose any concerted demands on the reconstruction.

**Table 2.** As in Table 1 except that the contributions to the curvature constraint integral from the two points of highest curvature are ignored, producing a kink at each of these points.

$\alpha$	$\Lambda = \frac{1}{2}\chi^2$	$\Pi = \alpha C$	$C$	$\hat{\gamma}$
250	900.8	4217.5	16.9	24
25	796.5	446.8	17.9	40
7.5	789.3	138.7	18.5	68
2.5	786.1	48.7	18.7	112
0.75	779.4	19.3	25.8	124

In our present context, a kink is allowed to occur at a particular vertex by setting the weight in the constraint calculation (7) to zero at that vertex. The effect of doing this at the two locations identified above is shown in the right-hand image of Fig. 5. The MAP solution for the same value of  $\alpha$  results in the desired appearance of uncorrelated backprojected noise. The reconstructions obtained under these conditions are shown in Fig. 6. Even for the highest value of  $\alpha$  of 250, the reconstruction matches the original object quite well, just because the kinks allow the boundary to yield better to the forces arising from the data. The match to the original object barely improves for lower  $\alpha$ . This observation is reflected in the fact that the values of the minus-log-likelihood ( $\Lambda$ ), given in Table 2, decrease very little as  $\alpha$  decreases. We also note that while the introduction of the kinks has not affected  $\hat{\gamma}$  very much,  $\alpha C$  has been reduced considerably. The result is that Gull’s condition for the appropriate strength of the prior is met at around  $\alpha = 5$ . The incorporation of kinks pushes the preferred  $\alpha$  value higher, favoring a smoother boundary.

The important issue is whether we can determine from the data that it is reasonable to add the kinks. The *AIC* requires that the minus-log-likelihood decrease by more than unity for each additional parameter that is introduced into a model. Adding a kink effectively increases the number of parameters by two, as can be seen by consideration of Bézier patches. A reduction of the constraint across a patch boundary from  $C^1$  to  $C^0$ , removes two constraint equations, one for each spatial variable, and hence adds two parameters that need to be determined. As can be seen by comparing Tables 1 and 2, the minus-log-likelihood always drops by at least four units and so adding these kinks is justified by the *AIC*.



**Figure 6.** As in Fig. 3, except that the contributions to the curvature constraint integral are removed at the two vertices of maximum curvature in that figure.

## 6. DISCUSSION

It should be clear that the Bayesian approach to analysis of data goes well beyond just estimating the parameters in a model that best fit a set of data. It allows one to make inferences about which models are more likely and provides guidance for extending models in the light of the data.

Although we used the *AIC* in this study to justify adding kinks to a boundary, a fully Bayesian approach would be preferable because it would allow one to explicitly incorporate prior knowledge into the decision process. After all, in a specific situation it might be well understood that kinks are extremely unlikely. One would therefore want to allow kinks only under extreme pressure from the data. It should also be said that the rules of probability require that one not really choose a specific value of a nuisance parameter like  $\alpha$ . Instead, one should marginalize (integrate) over  $\alpha$  in obtaining the MAP estimate of the model parameters. This process might be realizable, even for problems involving many parameters like this one, through the use of Markov Chain Monte Carlo (MCMC) methods<sup>20</sup> to randomly sample the full posterior  $p(\alpha, \mathbf{a}|\mathbf{d}, M_l)$ .

An interesting extension to this preliminary work is to consider the contribution to the curvature sum (7) at each vertex of the polygon model for the boundary to be a hyperparameter. Thus, instead of trying to determine  $\alpha$ , we would drop  $\alpha$  and try to determine every one of the weights  $w_j$  in Eq. (7). While this problem might seem intractable, it may be soluble using MCMC methods.

## ACKNOWLEDGEMENTS

This work was supported by the U.S. Dept. of Energy under Contract W-7405-ENG-36.

## 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, 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.
4. D. J. C. MacKay, "Bayesian interpolation," *Neural Computation* **4**, pp. 415–447, 1992.
5. S. F. Gull, "Developments in maximum-entropy data analysis," in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., pp. 53–71, Kluwer Academic, 1989.
6. H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *2nd Int. Symp. Information Theory*, pp. 267–281, Akademia Kaido, Budapest, 1973.
7. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, London, 1995.
8. S. F. Gull, "Bayesian inductive inference and maximum entropy," in *Maximum Entropy and Bayesian Methods in Science and Engineering (Vol. 1)*, G. J. Erickson and C. R. Smith, eds., pp. 53–74, Kluwer Academic, Dordrecht, 1989.
9. K. M. Hanson and G. S. Cunningham, "The hard truth," in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., Kluwer Academic, Dordrecht, 1994 (to be published).
10. 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.
11. J. Skilling, "The eigenvalues of mega-dimensional matrices," in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., pp. 455–466, Kluwer Academic, Dordrecht, 1989.
12. Y. Sakamoto, M. Ishiguro, and G. Kitagawa, *Akaike information criterion statistics*, D. Reidel, Dordrecht, 1986.
13. H. Akaike, "A Bayesian analysis of the minimum AIC procedure," *Ann Inst. Statist. Math.* **30**, pp. 9–14, 1978.
14. S. Lobregt and M. A. Viergever, "A discrete dynamic contour model," *IEEE Trans. Med. Imaging* **MI-14**, pp. 12–24, 1995.
15. 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).
16. 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.
17. 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.
18. 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.
19. G. S. Cunningham, I. Koyfman, and K. M. Hanson, "Improved convergence of gradient-based reconstructions using multi-scale models," to be published in *Image Processing*, M. H. Loew and K. M. Hanson, eds., *Proc. SPIE* **2710**, 1996.
20. J. Besag, P. Green, D. Higdon, and K. Mengersen, "Bayesian computation and stochastic systems," *Stat. Sci.* **10**, pp. 3–66, 1995.