

Simultaneous object estimation and image reconstruction in a Bayesian setting

K. M. Hanson

Los Alamos National Laboratory, MS P940
Los Alamos, NM 87545 USA

ABSTRACT

Suppose that it is desired to estimate certain parameters associated with a model of an object that is contained within a larger scene and that only indirect measurements of the scene are available. The optimal solution is provided by a Bayesian approach, which is founded on the posterior probability density distribution. The complete Bayesian procedure requires an integration of the posterior probability over all possible values of the image exterior to the local region being analyzed. In the present work, the full treatment is approximated by simultaneously estimating the reconstruction outside the local region and the parameters of the model within the local region that maximize the posterior probability. A Monte Carlo procedure is employed to evaluate the usefulness of the technique in a signal-known-exactly detection task in a noisy four-view tomographic reconstruction situation.

1. INTRODUCTION

Consider the problem of detecting a known object within a complex scene when given only indirect measurements, such as projections of the scene. A standard solution would be to reconstruct the full scene from the available data and then make the decision on the basis of how closely the reconstruction resembled the known object. Although this approach is straightforward, it might not yield optimal detection.

Here we seek a computational method that yields optimal performance of the stated detection task. The concept of such an 'ideal observer' has been useful in the past to help define the ultimate precision with which one can interpret data of a given type [1,2,3,4,5,6]. We propose a fully Bayesian approach in which the decision is based on the posterior probability. To achieve this solution, the reconstruction and the object parameters must be estimated simultaneously to obtain a fully self-consistent Bayesian decision.

Examples of this Bayesian estimation procedure are presented in a computed tomographic situation in which a nonnegativity constraint on the image is incorporated. The performance of the comprehensive Bayesian procedure is compared to that obtained by the traditional two-step approach using a Monte Carlo simulation of the entire imaging process [7,8].

This work is founded on the Bayesian concepts developed by Gull and Skilling and their colleagues [9,10,11,12], albeit under the assumption of a Gaussian distribution for the prior probability rather than their preferred entropic form.

2. THE BAYESIAN APPROACH

The foundation of reckoning in the Bayesian approach is the posterior probability, which is assumed to summarize the full state of knowledge concerning a given situation. The posterior probability is assigned the ultimate responsibility in making decisions about any hypotheses. Given the data set D , the posterior probability of any hypothesis \mathcal{H}_i is given by Bayes' law in terms of the proportionality

$$P(\mathcal{H}_i|D) \propto P(D|\mathcal{H}_i) P(\mathcal{H}_i), \quad (1)$$

where $P(D|\mathcal{H}_i)$, the probability of the observed data given hypothesis \mathcal{H}_i , is called the likelihood and $P(\mathcal{H}_i)$ is the prior probability of hypothesis \mathcal{H}_i . The likelihood is specified by the assumed probability distribution of the fluctuations in the measurements about their predicted values. The prior probability $P(\mathcal{H}_i)$ encompasses the full prior information about the relative frequency of occurrences of all hypotheses. Any known constraints concerning impossible hypotheses ought to be included explicitly or implicitly in $P(\mathcal{H}_i)$.

Making a binary decision is the simplest possible type of hypothesis testing, because there are just two alternative models between which to choose. The best decision variable is the ratio of posterior probabilities:

$$\frac{P(\mathcal{H}_1|D)}{P(\mathcal{H}_2|D)} = \frac{P(D|\mathcal{H}_1)P(\mathcal{H}_1)}{P(D|\mathcal{H}_2)P(\mathcal{H}_2)}. \quad (2)$$

Inherent in Bayesian analysis is the concept of cost functions, which state the relative importance of making correct versus incorrect decisions for each state of truth. The above decision strategy codified by Eq. (2) holds in the absence of asymmetric costs. When a continuum of possible outcomes exists, as in the estimation of one (or many) continuous parameters, the best possible choice of parameter values depends upon the type of cost function that is appropriate. It may be argued that for most general analyses, the simplest rule is to find the parameters that maximize the posterior probability, which is called the maximum *a posteriori* (MAP) solution [13,14].

We note that for an image containing N pixels, the MAP solution (or a reconstruction of any type) corresponds to a single point in an N -dimensional space. Any analysis based solely on such a reconstruction must necessarily ignore the complexity of the full posterior-probability distribution, which corresponds to a cloud in the same N -dimensional space. It is the correlations embodied in the posterior-probability distribution that we wish to incorporate in the present analysis.

In many problems there exist parameters that may be necessary to fully describe the solution, but whose values are of no interest. These unnecessary parameters can transform a simple hypothesis test into one of testing composite hypotheses. In such cases the proper approach is to integrate the probability density distribution over these unwanted variables. The result of this integration is called the marginal probability, which simply means summing over the irrelevant parameters in the problem (the result presumably to be written in the margin of the tally sheet).

2.1 Posterior Probability

We assume that there exists a scene that can be adequately represented by an orderly array of N pixels. We are given M discrete measurements that are linearly related to the original image amplitudes. We assume that these measurements are degraded by additive noise with a known covariance matrix \mathbf{R}_n , which describes the correlations that exist between noise fluctuations. The measurements can then be represented by a vector of length M

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \quad (3)$$

where \mathbf{f} is the original image vector of length N , \mathbf{n} is the random noise vector, and \mathbf{H} is the measurement matrix. In computed tomography the j th row of \mathbf{H} describes the weight of the contribution of image pixels to the j th projection measurement.

Now, because the probability is a function of continuous parameters, namely the N pixel values of the image, it is actually a probability density, designated by a small $p()$. The negative logarithm of the posterior probability is given by

$$-\log[p(\mathbf{f}|\mathbf{g})] = \phi(\mathbf{f}) = \Lambda(\mathbf{f}) + \Pi(\mathbf{f}), \quad (4)$$

where the first term comes from the likelihood and the second term from the prior probability. For additive Gaussian noise, the negative log(likelihood) is just half of chi-squared

$$-\log[p(\mathbf{g}|\mathbf{f})] = \Lambda(\mathbf{f}) = \frac{1}{2}\chi^2 = \frac{1}{2}(\mathbf{g} - \mathbf{H}\mathbf{f})^T \mathbf{R}_n^{-1}(\mathbf{g} - \mathbf{H}\mathbf{f}), \quad (5)$$

which is quadratic in the residuals. Instead of a Gaussian distribution assumed here, the Poisson distribution is often a better model for expected measurement fluctuations. The choice should be based on the statistical characteristics of the measurement noise, which we assume are known *a priori*.

The second term $\Pi(\mathbf{f})$ comes from the prior-probability distribution. It should incorporate as much as possible the known characteristics of the original image. Here we use a Gaussian distribution for the prior, whose negative logarithm may be written as

$$-\log[p(\mathbf{f})] = \Pi(\mathbf{f}) = \frac{1}{2}(\mathbf{f} - \bar{\mathbf{f}})^T \mathbf{R}_f^{-1}(\mathbf{f} - \bar{\mathbf{f}}), \quad (6)$$

where $\bar{\mathbf{f}}$ is the mean and \mathbf{R}_f is the covariance matrix of the prior-probability distribution. As we have done before [15], we invoke the prior knowledge that the image \mathbf{f} cannot have any negative components by instituting nonnegativity as a side constraint.

Another choice for prior that has been argued [11] to play a unique role for additive positive distributions is that of entropy. The Bayesian approach does not hinge on any particular choice of prior. However, the prior influences the outcome of the Bayesian procedure. For example, the strength of the prior affects the amount the reconstruction is biased away from the true image [16,17]. It is important for the researcher to understand the characteristics of solutions obtained regardless of the prior chosen. It is recognized that the prior provides the regularization essential to solving ill-posed problems [18,19], which arise because \mathbf{H} possesses a null-space [20,21].

2.2 Reconstruction Problem

In the reconstruction problem, we seek to estimate all pixel values in the original scene. The Bayesian solution to this problem is the image that maximizes the posterior probability or, equivalently, minimizes $-\log(\text{posterior probability})$. For the unconstrained MAP solution $\hat{\mathbf{f}}$, it is necessary that

$$\nabla_{\mathbf{f}} \phi = \mathbf{R}_f^{-1}(\mathbf{f} - \bar{\mathbf{f}}) + \mathbf{H}^T \mathbf{R}_n^{-1}(\mathbf{g} - \mathbf{H}\mathbf{f}) = 0. \quad (7)$$

However, under the constraint that the solution should be nonnegative ($f_i \geq 0$), Eq. (7) needs to be satisfied only when $f_i > 0$; a negative gradient is permissible on the boundary $f_i = 0$. In computed tomography (CT), the matrix operation \mathbf{H}^T is the familiar backprojection process.

A consequence of the prior is to bias the reconstruction away from the actual value in the original image, which was studied by Hanson [16] in unconstrained tomographic reconstructions. The extent of the bias depends on the relative weights of the two terms in Eq. (4). As the prior contribution vanishes the result approaches the maximum likelihood (or least-square residual) solution.

2.3 Analysis of a Local Region

Suppose that we ask a different question: does an object exist at a specific location in the image? The rest of the image is to be ignored. To address this question, we assume that within the image domain, a local region \mathcal{D} is to be analyzed for the possible presence of an object. Inside \mathcal{D} the image \mathbf{f} is assumed to be given by a model of the object described by a set of parameters α . Now the parameters in the problem are not the full set of image values \mathbf{f} , but rather $\mathbf{f}_{\mathcal{E}}$, the image values in the disjoint exterior region \mathcal{E} , and α . With Bayes' law the posterior probability may be written as $p(\mathbf{f}_{\mathcal{E}}, \alpha | \mathbf{g}) \propto p(\mathbf{g} | \mathbf{f}_{\mathcal{E}}, \alpha) p(\mathbf{f}_{\mathcal{E}}, \alpha) \propto p(\mathbf{g} | \mathbf{f}_{\mathcal{E}}, \alpha) p(\mathbf{f})$. In the last step we have chosen to avoid explicit specification of a prior on α , allowing it to be implicitly included in the prior for \mathbf{f} .

As the new question regards only the region \mathcal{D} , the image values $\mathbf{f}_{\mathcal{E}}$ outside \mathcal{D} are irrelevant. The Bayesian approach specifies that we integrate the posterior probabilities over the unwanted parameters of the problem, namely over the image values outside \mathcal{D} . If the problem at hand is to decide between two possible sets of parameters, α_1 or α_2 , the decision variable should be the ratio of the two marginal posterior probabilities [13,14], or equivalently its logarithm

$$\psi = \log \left[\frac{\int_{\mathcal{E}} p(\mathbf{f}_{\mathcal{E}}, \alpha_1 | \mathbf{g}) d\mathbf{f}}{\int_{\mathcal{E}} p(\mathbf{f}_{\mathcal{E}}, \alpha_2 | \mathbf{g}) d\mathbf{f}} \right], \quad (8)$$

where the integrals are to be carried out only over the external region \mathcal{E} and include all possible image values not disallowed by constraints. Within the context of Bayesian analysis, this decision variable logically follows from the statement of the problem. Hence, we assert that it should yield optimal decisions. The ideal observer uses Eq. (8) to make binary decisions regarding a local region.

Under certain circumstances these integrals may be difficult to calculate accurately. However, when dealing with the Gaussian prior- and likelihood-probability density distributions presented in Sec. 2.1, we expect the posterior-probability density $p(\mathbf{f}_{\mathcal{E}}, \boldsymbol{\alpha}|\mathbf{g})$ to decrease rapidly away from a unique maximum. Using $\hat{\mathbf{f}}_{\mathcal{E}k}$ to designate the image in the exterior region that maximizes the posterior probability for the parameter set $\boldsymbol{\alpha}_k$, we are prompted to rewrite the above ratio as,

$$\psi = \log \left[\frac{p(\hat{\mathbf{f}}_{\mathcal{E}1}, \boldsymbol{\alpha}_1|\mathbf{g}) K(\boldsymbol{\alpha}_1)}{p(\hat{\mathbf{f}}_{\mathcal{E}2}, \boldsymbol{\alpha}_2|\mathbf{g}) K(\boldsymbol{\alpha}_2)} \right], \quad (9)$$

where the phase-space factor is

$$K(\boldsymbol{\alpha}_k) = \frac{1}{p(\hat{\mathbf{f}}_{\mathcal{E}k}, \boldsymbol{\alpha}_k|\mathbf{g})} \int_{\mathcal{E}} p(\mathbf{f}_{\mathcal{E}}, \boldsymbol{\alpha}_k|\mathbf{g}) d\mathbf{f}, \quad (10)$$

which accounts for the extent of the spread in \mathbf{f} -space of the posterior-probability density distribution about its constrained peak value $p(\hat{\mathbf{f}}_{\mathcal{E}k}, \boldsymbol{\alpha}_k|\mathbf{g})$. Generally $\hat{\mathbf{f}}_{\mathcal{E}1} \neq \hat{\mathbf{f}}_{\mathcal{E}2}$ because a change in model parameters describing the interior region alters the projections, implying that a different exterior image will minimize the posterior probability. In many situations, however, replacing the local region of the MAP solution with the model may have little effect on the predicted projection values. Then, $p(\mathbf{f}_{\mathcal{E}}, \boldsymbol{\alpha}|\mathbf{g})$ is independent of $\boldsymbol{\alpha}$ and, to good approximation, $\hat{\mathbf{f}}_{\mathcal{E}1} = \hat{\mathbf{f}}_{\mathcal{E}2} = \hat{\mathbf{f}}_{\mathcal{E}}$, so both K factors in Eq. (9) are the same and

$$\psi = \log \left[\frac{p(\hat{\mathbf{f}}_{\mathcal{E}}, \boldsymbol{\alpha}_1|\mathbf{g})}{p(\hat{\mathbf{f}}_{\mathcal{E}}, \boldsymbol{\alpha}_2|\mathbf{g})} \right]. \quad (11)$$

In these situations, the decision variable can be given adequately by the change in the log(posterior probability) induced by replacing the MAP solution $\hat{\mathbf{f}}$ in \mathcal{D} with the two models, leaving the exterior region unchanged.

For unconstrained solutions of Eq. (7), the K factor is independent of $\boldsymbol{\alpha}$, because the shape of the Gaussian posterior-probability distribution is governed by the full curvature of ϕ , namely $\mathbf{R}_{\mathbf{n}}^{-1} + \mathbf{R}_{\mathbf{f}}^{-1}$. Then the K factors in Eq. (9) cancel and

$$\psi = \log \left[\frac{p(\hat{\mathbf{f}}_{\mathcal{E}1}, \boldsymbol{\alpha}_1|\mathbf{g})}{p(\hat{\mathbf{f}}_{\mathcal{E}2}, \boldsymbol{\alpha}_2|\mathbf{g})} \right]. \quad (12)$$

The argument of the logarithm is called the generalized posterior-probability ratio. Equation (12) may not be a good approximation to (8) for constrained solutions, as the contribution to the phase-space K factor from the integral over each \mathbf{f}_i depends on the relation of the peak in $\hat{\mathbf{f}}_{\mathcal{E}i}$ to the constraint boundary. Nonetheless, because of its simplicity, we use Eq. (12) and reserve for the future an investigation of a better approximation.

To evaluate Eq. (12) for fixed parameter sets $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$, it is necessary to find the pair of exterior images, $\mathbf{f}_{\mathcal{E}1}$ and $\mathbf{f}_{\mathcal{E}2}$, that maximize the posterior-probability density. In other words, one must find the maximum *a posteriori* or MAP reconstruction in the exterior region with the image inside the local region fixed by the parameter values. To extend the binary decision problem to one in which the model parameters are to be estimated, it becomes necessary to simultaneously estimate the $\boldsymbol{\alpha}$ parameters and reconstruct the exterior region with the aim of minimizing the posterior probability.

2.4 Method of Solution

We employ the same iterative method described in [15] to find the constrained MAP solutions. Following Butler, Reeds, and Dawson [22], the constrained solution is the positive part of the dual functional \mathbf{d} :

$$\mathbf{f}_i^k = \mathbf{d}_i^k, \quad \mathbf{d}_i^k \geq 0 \quad (13)$$

$$\mathbf{f}_i^k = 0, \quad \mathbf{d}_i^k < 0. \quad (14)$$

The first estimate is $\mathbf{d}^0 = \bar{\mathbf{f}}$. The gradient of ϕ for the k th estimate,

$$\nabla_{\mathbf{f}} \phi^k = \mathbf{r}^k = \mathbf{R}_{\mathbf{f}}^{-1}(\mathbf{f}^k - \bar{\mathbf{f}}) + \mathbf{H}^T \mathbf{R}_n^{-1}(\mathbf{g} - \mathbf{H}\mathbf{f}^k) \quad (15)$$

is used to update \mathbf{d}^k :

$$\mathbf{d}^{k+1} = \mathbf{d}^k + c^k \mathbf{r}^k, \quad (16)$$

where scalar c^k is chosen to minimize ϕ^{k+1} . This optimization procedure is essentially that of steepest descent, which is known to be inefficient. Although a more rapid solution method probably exists, the present one suffices for this study. It is found that 50 iterations are adequate for the present work.

To fit the parameters of the model in the local region, we also need the gradient of ϕ with respect to the parameters for positive (negative) increments of α_j

$$\frac{\partial \phi^{k+(-)}}{\partial \alpha_j} = \sum_i \frac{\partial \phi^k}{\partial f_i^k} \frac{\partial f_i^k}{\partial \alpha_j}, \quad (17)$$

where the sum excludes pixels for which either $d_i^k < 0$ or $d_i^k = 0$, when the product of the partial derivatives is positive for (+) or negative for (-). These contributions are excluded because they would move d_i^k further into the domain of forbidden solutions and hence produce no change in f_i^k .

3. METHOD

We demonstrate the use of the Bayesian approach in making decisions about a local region in a reconstructed image with a very simple example: detection of disks based on a very limited number of noisy projections. This binary discrimination task is employed because it is theoretically tractable, it is easy to perform the required decision-making procedure, and it is possible to summarize the results simply.

3.1 Monte Carlo Method to Evaluate Task Performance

The overall method for evaluating a reconstruction algorithm used here has been described before [7,8]. In this method a task performance index for a specified imaging situation is numerically evaluated. The technique is based on a Monte Carlo simulation of the entire imaging process including random scene generation, data taking, reconstruction, and performance of the specified task. The accuracy of the task performance is determined by comparison of the results with the known original scene using an appropriate figure of merit. Repetition of this process for many randomly generated scenes provides a statistically significant estimate of the performance index [8].

3.2 Specifications of Detection Tests

The imaging situation is chosen in an attempt to maximize the possible effect of re-estimation of the exterior region implied by the full Bayesian treatment. The original scenes contain either one or two disks, all with amplitude 0.1 and diameter 8 pixels. The disks are randomly placed, but not overlapping, within the circle of reconstruction of diameter 64 pixels. The background level is zero. Enough scenes are generated in the testing sequence to provide 100 disks with amplitude 0.1 and 100 null disks to sample the background region.

The measurements consist of four parallel projections, each containing 64 samples, taken at 45° increments in view angle. Measurement noise is simulated by adding to each measurement a pseudorandom number taken from a Gaussian distribution with a standard deviation of 2. The peak projection value of one of the disks is 0.80. The signal-to-noise ratio (SNR) for signal-known-exactly (SKE) detection of a disk may be easily calculated as the $[\sum \text{SNR}_i^2]^{1/2}$, summed over the measurements that subtend the disk, which yields $\text{SNR}_{\text{detect}} = d' = 1.89$. The overwhelming difficulty of this detection problem can be seen in Fig. 1, which shows the first projection of the first scene. To avoid aliasing artifacts in the reconstruction, the projection data used for reconstruction are presmoothed using a triangular convolution kernel with a FWHM of 3 samples. As a result, the expected rms noise value in the smoothed data is approximately

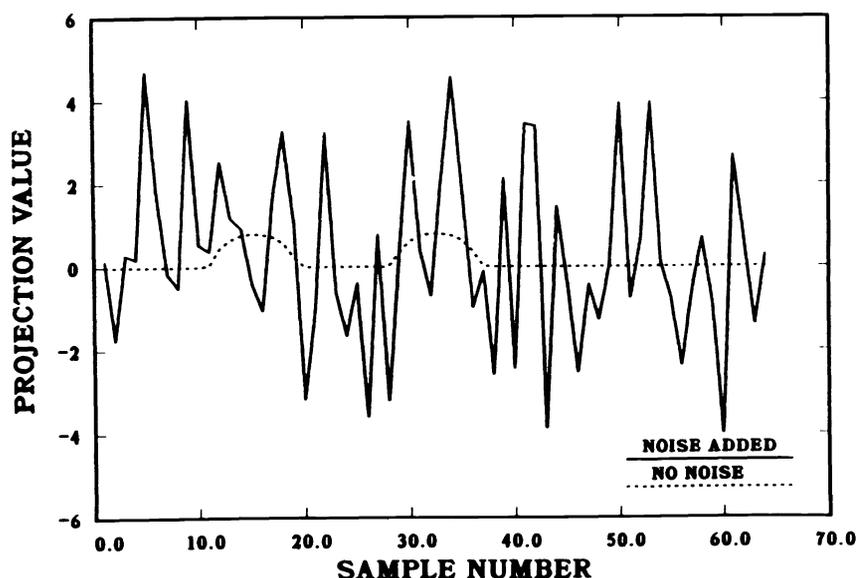


Fig. 1. The measurements that comprise the first of four projections of the first scene containing two disks.

1.0. Thus for all cases studied we use the noise covariance matrix $\mathbf{R}_n = \text{diag}(\sigma_n^2) = (1.0)^2$. With this assumption we are ignoring the correlations in the data caused by presmoothing.

For the Gaussian prior probability distribution we employ the ensemble mean $\bar{f}_i = 0.0031 = \text{constant}$, which is the average value of the scenes containing two disks. We assume the ensemble covariance matrix is diagonal with $\mathbf{R}_f = \text{diag}(\sigma_f^2)$ and explore the effect of choosing different values of σ_f .

The stated task is to detect the presence of the disks under the assumption that the signal is known exactly (SKE) and the background is known exactly (BKE) in the 2D local region. The various strategies for making this binary decision are presented in the next section. A useful measure to summarize the performance of binary decisions is the detection index d_A , which is based on the area under the Receiver Operating Characteristic (ROC) curve. The ROC curve is obtained in the usual way [8] from the histograms in the decision variable for the signal-known-present and the signal-known-absent tests. Once the ROC curve is generated and its area A determined, then d_A is found by $d_A = 2 \text{erfc}^{-1}\{2(1 - A)\}$, where erfc^{-1} is the inverse complement of the error function. There are good reasons for not using the detectability index d' , which is based on the first and second moments of the histograms of the decision variable [23]. For a fixed number of binary tests, the relative statistical error in d_A is smallest when d_A is about 2.2 [8]. The imaging situation should be arranged to keep d_A roughly between 1 and 3.5 to optimize the statistical value of the testing procedure.

3.3 Decision Strategies

For the simple binary discrimination tests performed here, only two parameters are needed to describe the model for the local region – the background level and the disk amplitude relative to the background. The background is assumed to be constant. The position and diameter of the disk are assumed to be known. The edge of the disk is linearly ramped over 2 pixels in radius to roughly match the blur caused by the reconstruction process. The local region of analysis is assumed to be circular with a diameter of 14 pixels and centered on the test position. When the disk is assumed present, the amplitude is set to 0.1 and when assumed absent, 0. The background level is 0 for both tests. In all of the decision strategies, a decision variable is evaluated for each of the two hypotheses and the difference between the two values is used to make the decision whether a disk is present or not.

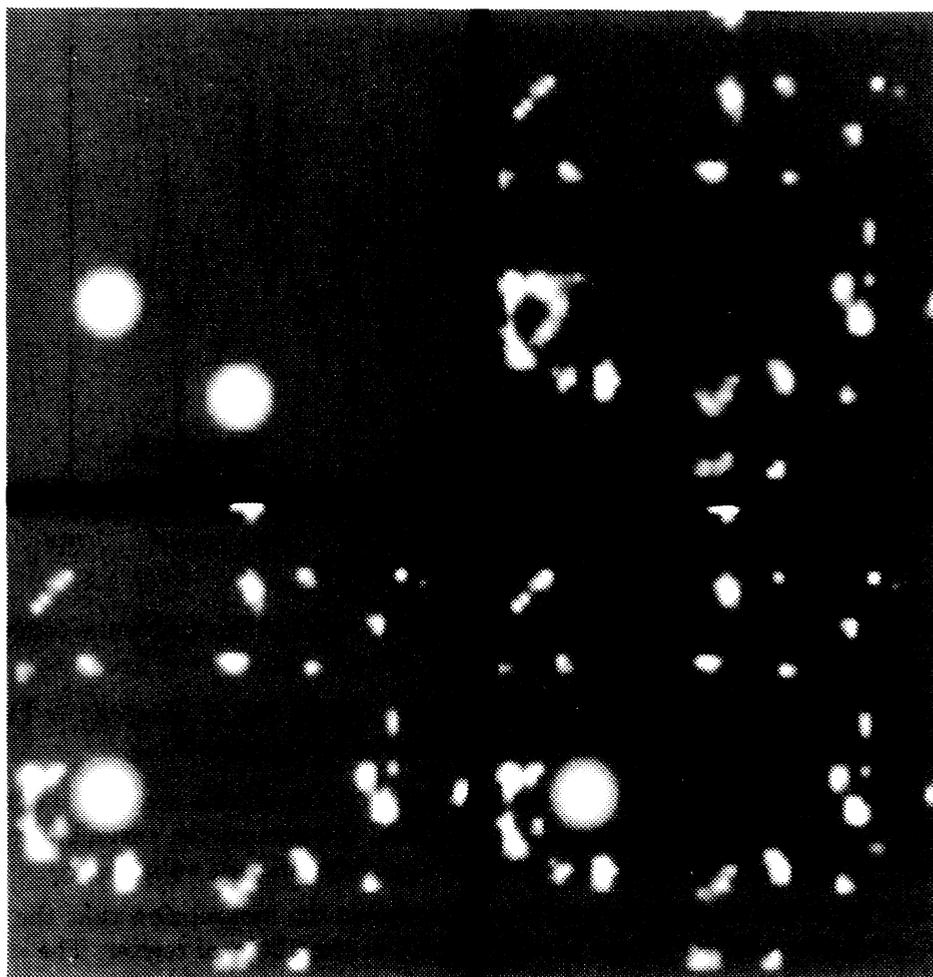


Fig. 2. This composite image shows the process used to make the binary decision whether a disk is present or not. The original scene (upper left) is reconstructed from four projections using constrained maximum *a posteriori* reconstruction (upper right) with ensemble standard deviation $\sigma_f = 1$. To test the possible presence of a disk, that disk is placed into the reconstruction (lower left) and then the image outside the local region of the disk is 're-reconstructed' to obtain the image (lower right) that maximizes the posterior probability with the disk present. This procedure is repeated with the same region replaced by the background value (zero) and the difference in the logarithms of the two posterior probabilities is used as the decision variable.

The following decision strategies are employed in this study:

Method A) In the simplest possible approach, one uses the projection data directly. The decision is based on the difference in χ^2 for the two hypotheses. Explicitly, Eq. (5) is evaluated under both hypotheses using for \mathbf{f} only the model values inside the local region of analysis \mathcal{D} . The image values outside the analysis region are implicitly assumed to be zero. If the background is truly zero and only one disk is present in the scene, this decision variable operates at the statistical limit attainable in the absence of prior information. However, it is obviously deficient for complex scenes as it ignores the contributions to the projections arising from features outside the local region.

Method B) By Bayesian reckoning, the best possible decision variable for local analysis is given by Eq. (8). For this method we use the approximation given by the generalized posterior-probability ratio Eq. (12), which implies that for each choice of parameters the exterior region is reconstructed to maximize $p(\mathbf{f}_\varepsilon, \boldsymbol{\alpha} | \mathbf{g})$.

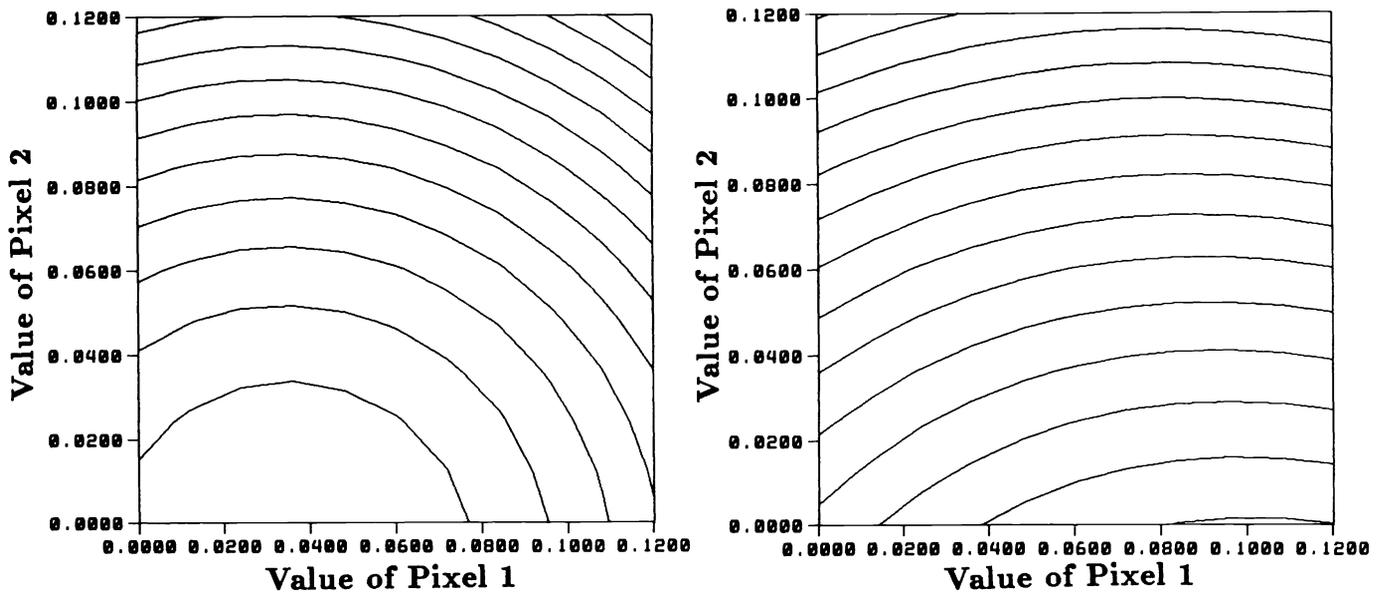


Fig. 3. Contour plots showing the correlation in $-\log(\text{posterior probability})$ for fluctuations in two pixel values about the MAP solution. These results are shown for two different assumed values of the ensemble standard deviation σ_f ; 0.2 (left) and 1.0 (right) for which the contour spacing is 0.01 and 0.05, respectively. The first pixel is centered on the lower middle disk in the first scene (Fig. 2) and the other is three pixels down and three pixels to the left of the first.

In actual practice, this second reconstruction step follows a preliminary constrained MAP reconstruction of the whole image as pictorially described in Fig. 2.

Method C) This method uses Eq. (11) for the decision variable based on the posterior-probability distribution associated with the MAP reconstruction. Readjustment of the reconstruction external to the analysis region for each test hypothesis is not required. This method was introduced by Gull and Skilling [12] and studied by Myers and Hanson [17] for an entropy prior.

Method D) Method D proceeds from the constrained MAP reconstruction $\hat{\mathbf{f}}_{\text{MAP}}$ from the data. The decision variable is taken as the difference in $|\mathbf{f} - \hat{\mathbf{f}}_{\text{MAP}}|^2$ for the two models hypothesized for the local region. This method was used by Hanson and Myers [15] to compare performance of the Rayleigh task using MAP reconstructions based on Gaussian and entropy priors. It corresponds to using a likelihood approach based on the reconstruction in which the noise fluctuations in the reconstruction are assumed to be uncorrelated and Gaussian distributed. This method therefore ignores the correlations in the posterior-probability distribution, shown in Fig. 3, that are incorporated to various degrees by methods B and C.

Method E) Method E also proceeds from the constrained MAP reconstruction $\hat{\mathbf{f}}_{\text{MAP}}$. Unlike the preceding methods, the amplitude and background are varied to find the combination of values that minimizes $|\mathbf{f} - \hat{\mathbf{f}}_{\text{MAP}}|^2$. In this fitting process, both the relative amplitude and the background are constrained to be nonnegative. The amplitude so determined is used as the decision variable. This method was used by Hanson in many earlier studies [7,24,25,26,8,16]. It is closely related to the non-prewhitening matched filter, which is optimal when the fluctuations in the reconstruction are uncorrelated and Gaussian distributed.

A human observer viewing a reconstruction does not have access to the posterior probability distribution and thus may have to resort to employing a decision method similar to D or E.

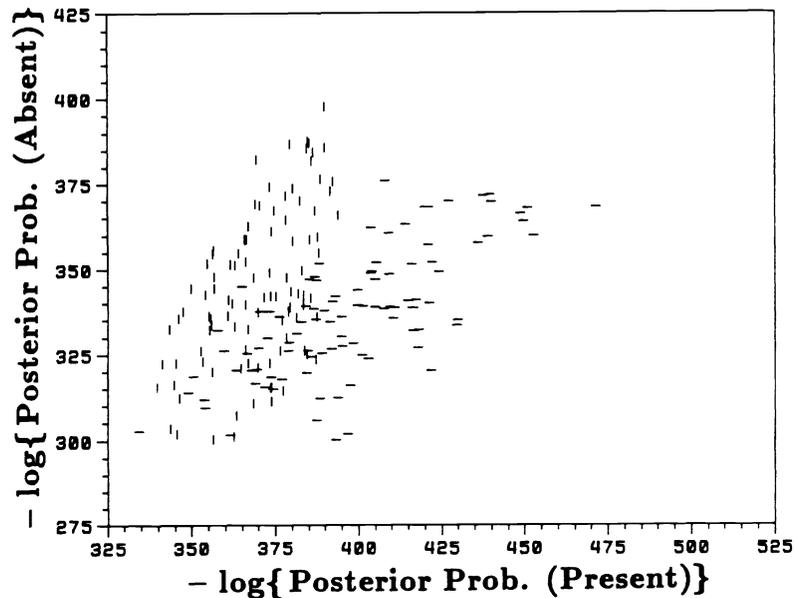


Fig. 4. Scatter plot showing $-\log(\text{posterior probability})$ calculated using Eq. (12) for the two separate hypotheses; (1), a disk is present and (2), no disk is present. The vertical bars correspond to tests made at sites where disks are actually present and horizontal bars where no disk is present. The degree of separation between the two clouds indicates the ability to detect the disks. The difference between the ordinate and abscissa values is the decision variable for method B. This result is for two disks per scene and $\sigma_f = 0.2$.

4. RESULTS

A constrained MAP reconstruction of the first scene of the testing sequence for two disks is shown in Fig. 2. Because of the noise in the projection data, the presence of the disks in the original scene is obscured in the reconstruction. An interesting aspect of the posterior-probability approach is that one may calculate the probability of a disk being present at any location in the reconstruction. Even though the reconstruction might be zero (the lower limit decreed by the constraint of nonnegativity) throughout a certain region, the probability of a disk being present in that region is finite and calculable. By contrast, any analysis method based solely on the reconstruction would not be able to distinguish two different regions that are completely zero. This point is emphasized by the contour plots in Fig. 3, which show the behavior of the posterior-probability distribution when the values of two nearby pixels are varied. In both cases the MAP solution for one of the pixels is zero, although both pixels actually fall within a disk in the original scene and should have the value 0.1.

Figure 4 shows the distributions of the two components of the decision variable used in method B obtained for the two hypotheses (disk present and disk absent). The two clouds formed for each state of truth have an approximate Gaussian shape. The decision variable for this method is the difference between the ordinate and abscissa values.

The test sequences generated to demonstrate the use of posterior probability in decision making are analyzed for several different values of the ensemble covariance matrix σ_f . As we have found before [15,16,25,26,17], the performance of visual-like tasks usually varies with the parameters that control the rms residual achieved by the reconstruction algorithm. For the present MAP algorithm, that parameter is the ratio σ_f/σ_n . Recall that σ_n is fixed at its expected value of 1.0. For $\sigma_f = 0.1, 0.2,$ and $1.0,$ the rms residuals of the constrained MAP reconstructions of the single-disk scenes are 0.84, 0.80, and 0.76, respectively. As expected, the bias in the reconstructions also depends on σ_f . The disk amplitudes, measured as the average value over each disk relative to the average over its surrounding annulus (essentially method E), are 0.007, 0.014, and 0.025, for the same σ_f sequence. These are far from the actual value of

Table 1. Summary of detectabilities obtained using different methods to analyze an identical data set derived from 100 separate scenes each containing a single randomly placed disk. These results are presented for rms noise $\sigma_n = 1$, but for different assumed values of σ_f .

Method	Decision variable	d_A		
		$\sigma_f = 0.1$	$\sigma_f = 0.2$	$\sigma_f = 1$
A	$\Delta\chi^2$ (use data only)	1.98	same	same
B	$\Delta\log(\text{posterior probability})$ (exterior re-estimated)	2.11	2.07	1.95
C	$\Delta\log(\text{posterior probability})$ (exterior fixed at $\hat{\mathbf{f}}_{\text{MAP}}$)	2.11	2.04	1.95
D	$\Delta \mathbf{f} - \hat{\mathbf{f}}_{\text{MAP}} ^2$ (use reconstruction only)	1.88	1.73	1.57

Table 2. Summary of detectabilities obtained by analyzing data derived from 50 separate scenes with two disks placed randomly in each scene. Assumed are $\sigma_n = 1$ and a variety of σ_f values.

Method	Decision variable	d_A			
		$\sigma_f = 0.02$	$\sigma_f = 0.1$	$\sigma_f = 0.2$	$\sigma_f = 1$
A	$\Delta\chi^2$ (use data only)	1.75	same	same	same
B	$\Delta\log(\text{posterior probability})$ (exterior re-estimated)	1.80	1.87	1.82	1.74
C	$\Delta\log(\text{posterior probability})$ (exterior fixed at $\hat{\mathbf{f}}_{\text{MAP}}$)	1.81	1.87	1.81	1.70
D	$\Delta \mathbf{f} - \hat{\mathbf{f}}_{\text{MAP}} ^2$ (use reconstruction only)	1.80	1.76	1.67	1.47
E	$\Delta(\text{disk amplitude})$ (constrained fit to $ \mathbf{f} - \hat{\mathbf{f}}_{\text{MAP}} ^2$)	1.01	1.09	1.01	0.96

0.10, probably because there are so few views, giving rise to a gigantic null space [21], together with so much noise. The MAP algorithm based on a Gaussian prior with $\hat{\mathbf{f}}_i = 0$, which is nearly the case here, amounts to using minimum-norm regularization. Therefore, control of the noise, which dominates the reconstructed field, can only be effected by reducing the sensitivity of the reconstruction.

Tables 1 and 2 summarize the detectability results obtained in the tests described above. The absolute statistical accuracy of these d_A values is about 0.25. The d_A value for the single-disk scenes based on using just the measurement data (Method A) is 1.98, in good agreement with the value of 1.89 estimated in Sec. 3.2. As only the prior is involved, this value is independent of σ_f . The d_A value obtained with the same method in the two-disk tests is 1.75, but the decrement relative to the single-disk tests may not be statistically significant because the two test series involve different randomized data sets. Much better accuracy should prevail in comparisons within each table, however, because they are obtained by analyzing the exact same data sequence. We observe similar trends in both tables to remarkable accuracy. First of all, both methods of using the posterior probability (methods B and C) provide nearly the same detectability over a large range of σ_f values. Perhaps this consistent behavior stems from the ability of the posterior probability to fully retain the available information even though σ_f changes. There seems to be little

advantage to re-estimation of the exterior of the local region to minimize the posterior probability implied by Eq. (12) in this imaging situation. There is a trend toward better detectability as σ_f get smaller. The force of regularization imposed by the prior is overwhelming at $\sigma_f = .02$. For example, the average disk amplitude there is 0.0005 and the reconstruction values lie between 0.0005 and 0.0046; the nonnegativity constraint is not even engaged.

The methods based on the posterior probability yield slightly better ($\approx 10\%$) detectabilities than method D, which is based only on the reconstruction under the assumption that the noise uncertainty is uncorrelated and stationary. Basing the decision on the estimated disk amplitude (method E) greatly ($\approx 45\%$) reduces detectability compared to the other methods.

For **unconstrained** MAP with $\sigma_f = 0.1$, the d_A values for the single-disk test series are 2.10 (method B) and 2.08 (method C). These values are nearly the same as those in Table 1, so the nonnegativity constraint has little effect on detectability in the present situation. In previous work involving a limited number of views, we have seen remarkable improvements in detectability wrought by the nonnegativity constraint [7,24,25,8]. Although the less efficient method E was used in those studies, the principle reason for the ineffectiveness of nonnegativity in the present case is that it is more limited by noise than by the null space. The large amount of noise is used here to limit d_A within the range of reasonable accuracy as discussed in Sec. 3.2. The effects of artifacts were enhanced in previous studies by adding several disks with large amplitude to the scene.

The d_A value obtained from unconstrained MAP reconstructions ($\sigma_f = 0.1$) by method D is 2.07. The sizable gain in detectability over the result presented in Table 1 for constrained reconstructions, 1.88, may be explained by the close relationship between method D and a likelihood approach applied to the reconstruction. Such an approach is more likely to be valid when the reconstruction is not constrained.

5. CONCLUSION

We have compared several methods for detecting small disks in tomographic reconstructions. The worst performance is provided by method E in which the amplitude obtained by fitting the MAP reconstruction is used as the decision variable. This choice is the same as the matched filter for uncorrelated, Gaussian distributed noise fluctuations, so it is more appropriate for unconstrained reconstructions than for constrained reconstructions. A better decision variable is the mean-square difference between the model and the reconstruction $|\mathbf{f} - \hat{\mathbf{f}}_{\text{MAP}}|^2$, as it is similar to a log(likelihood ratio), again ignoring correlations in the reconstruction fluctuations. This method provides much better results. The best detectabilities is achieved by basing decisions on the calculated posterior probability, which takes fully into account the information contained in the measurements as well as in the prior knowledge. In the present tests, however, there seems to be little benefit in re-estimating the exterior region. The full Bayesian treatment codified by Eq. (8) is expected to represent the ideal observer.

ACKNOWLEDGEMENTS

I am indebted to Stephen F. Gull and John Skilling for their inciteful MEMSYS 3 Manual [12] and other publications, which substantially motivated this work. I wish to express my appreciation to Kyle J. Myers and Robert F. Wagner for many stimulating discussions and worthwhile suggestions. This work was supported by the United States Department of Energy under contract number W-7405-ENG-36.

References

- [1] K. M. Hanson. On the optimality of the filtered backprojection algorithm. *J. Comput. Assist. Tomogr.*, 4:361–363, 1980.
- [2] K. M. Hanson. Variations in task and the ideal observer. *Proc. SPIE*, 419:60–67, 1983.
- [3] R. F. Wagner, K. J. Myers, G. D. Brown, M. J. Tapiovaara, and A. E. Burgess. Higher order tasks: human vs. machine performance. *Proc. SPIE*, 1090:183–194, 1989.
- [4] A. E. Burgess and H. Ghandeharian. Visual signal detection. I. Ability to use phase information. *J. Opt. Soc. Amer.*, A1:900–905, 1984.

- [5] A. E. Burgess and H. Ghandeharian. Visual signal detection. II. Signal-location identification. *J. Opt. Soc. Amer.*, A1:906–910, 1984.
- [6] A. E. Burgess. Visual signal detection. III. On Bayesian use of prior information and cross correlation. *J. Opt. Soc. Amer.*, A2:1498–1507, 1985.
- [7] K. M. Hanson. Method to evaluate image-recovery algorithms based on task performance. *Proc. SPIE*, 914:336–343, 1988.
- [8] K. M. Hanson. Method to evaluate image-recovery algorithms based on task performance. *J. Opt. Soc. Amer.*, A7:45–57, 1990.
- [9] S. F. Gull. Developments in maximum-entropy data analysis. In J. Skilling, editor, *Maximum Entropy and Bayesian Methods*, pages 53–71, Kluwer Academic, 1989.
- [10] S. F. Gull. Bayesian inductive inference and maximum entropy. In G. J. Erickson and C. R. Smith, editors, *Maximum Entropy and Bayesian Methods in Science and Engineering (Vol. 1)*, pages 53–74, Kluwer Academic, 1989.
- [11] J. Skilling. Classic maximum entropy. In J. Skilling, editor, *Maximum Entropy and Bayesian Methods*, pages 45–52, Kluwer Academic, 1989.
- [12] S. F. Gull and J. Skilling. *Quantified Maximum Entropy - MEMSYS 3 Users' Manual*. Maximum Entropy Data Consultants Ltd., Royston, England, 1989.
- [13] H. L. Van Trees. *Detection, Estimation, and Modulation Theory - Part I*. John Wiley and Sons, New York, 1968.
- [14] A. D. Whalen. *Detection of Signals*. Academic, New York, 1971.
- [15] K. M. Hanson and K. J. Myers. Rayleigh task performance as a method to evaluate image reconstruction algorithms. In W. T. Grandy and L. H. Schick, editors, *Maximum Entropy and Bayesian Methods*, Kluwer Academic, 1991.
- [16] K. M. Hanson. Object detection and amplitude estimation based on maximum *a posteriori* reconstructions. *Proc. SPIE*, 1231:164–175, 1990.
- [17] K. J. Myers and K. M. Hanson. Comparison of the algebraic reconstruction technique with the maximum entropy reconstruction technique for a variety of detection tasks. *Proc. SPIE*, 1231:176–187, 1990.
- [18] M. Z. Nashed. Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory. *IEEE Trans. Antennas Propagat.*, AP-29:220–231, 1981.
- [19] D. M. Titterton. General structure of regularization procedures in image reconstruction. *Astron. Astrophys.*, 144:381–387, 1985.
- [20] K. M. Hanson and G. W. Wecksung. Bayesian approach to limited-angle reconstruction in computed tomography. *J. Opt. Soc. Amer.*, 73:1501–1509, 1983.
- [21] K. M. Hanson. Bayesian and related methods in image reconstruction from incomplete data. In Henry Stark, editor, *Image Recovery: Theory and Application*, pages 79–125, Academic, Orlando, 1987.
- [22] J. P. Butler, J. A. Reeds, and S. V. Dawson. Estimating solutions for first kind integral equations with nonnegative constraints and optimal smoothing. *SIAM J. Numer. Anal.*, 18:381–397, 1981.
- [23] R. F. Wagner, K. J. Myers, M. J. Tapiovaara, G. D. Brown, and A. E. Burgess. Maximum a posteriori detection and figures of merit for detection under uncertainty. *Proc. SPIE*, 1231:195–204, 1990.
- [24] K. M. Hanson. POPART - Performance Optimized Algebraic Reconstruction Technique. *Proc. SPIE*, 1001:318–325, 1988.
- [25] K. M. Hanson. Optimization for object localization of the constrained algebraic reconstruction technique. *Proc. SPIE*, 1090:146–153, 1989.
- [26] K. M. Hanson. Optimization of the constrained algebraic reconstruction technique for a variety of visual tasks. In D. A. Ortendahl and J. Llacer, editors, *Proc. Information Processing in Medical Imaging*, pages 45–57, Wiley-Liss, New York, 1990.