

Probing the covariance matrix

Kenneth M. Hanson

*T-16, Nuclear Physics, Los Alamos National Laboratory,
Los Alamos, New Mexico, USA 087545
kmh@lanl.gov*

Abstract. By drawing an analogy between the logarithm of a probability distribution and a physical potential, it is natural to ask the question, “what is the effect of applying an external force on model parameters?” In Bayesian inference, parameters are frequently estimated as those that maximize the posterior, yielding the maximum a posteriori (MAP) solution, which corresponds to minimizing $\varphi = -\log(\text{posterior})$. The uncertainty in the estimated parameters is typically summarized by the covariance matrix for the posterior distribution, \mathbf{C} . I describe a novel approach to estimating specified elements of \mathbf{C} in which one adds to φ a term proportional to a force, \mathbf{f} , that is hypothetically applied to the parameters. After minimizing the augmented φ , the change in the parameters is proportional to $\mathbf{C}\mathbf{f}$. By selecting the appropriate force, the analyst can estimate the variance in a quantity of special interest, as well as its covariance relative to other quantities. This technique allows one to replace a stochastic MCMC calculation with a deterministic optimization procedure. The usefulness of this technique is demonstrated with a few simple examples, as well as a more complicated one, namely, the uncertainty in edge localization in a tomographic reconstruction of an object’s boundary from two projections.

Key Words: covariance matrix estimation, probability potential, posterior stiffness, external force, probing the posterior

INTRODUCTION

Bayesian inference for large nonlinear problems is often carried out through numerical models and calculation[1]. The inference process requires estimates of the parameter values, $\hat{\mathbf{a}}$, and their uncertainties. The uncertainties are related to the width of the posterior and typically characterized in terms of the covariance matrix \mathbf{C} . The maximum a posteriori (MAP) solution is frequently chosen as an estimate of the parameters because it is easier to find than the posterior mean.

Standard approaches to determining \mathbf{C} include: 1) sensitivity analysis, 2) functional analysis based on sensitivities[2], and 3) Markov chain Monte Carlo[3]. Each of these approaches has its advantages and disadvantages, depending on the nature of the problem, including factors such as the number of parameters, the number of measurements, and the cost of evaluating the forward model and its sensitivities.

By drawing an analogy between the logarithm of the posterior and a physical potential, it is natural to think about applying an external force to the model parameters to determine the stiffness of the MAP solution. The resulting algorithm provides a deterministic way to estimate selected elements of the covariance matrix[4, 5, 6, 7]. The usefulness of this new technique is demonstrated with examples ranging from simple to complicated.

PHYSICAL ANALOGY

Relationships between statistics and physics often provide a deeper understanding that can lead to new or improved algorithmic approaches to solving statistics problems. In statistical physics[8], probability distributions are often written as the exponential of a physical potential. Thus, in Bayesian analysis, it is natural to draw an analogy between $\varphi(\mathbf{a}) = -\log(p(\mathbf{a} | \mathbf{y}))$ and a physical potential, where $p(\mathbf{a} | \mathbf{y})$ is the posterior, the vector \mathbf{a} represents the n continuous parameters, and \mathbf{y} represents the m measurements.

The analogy between probabilities and potentials has been used in many applications, for example, to develop priors on deformable surfaces based on the mechanical properties of metallic rods[7]. It has been used to develop novel Markov Chain Monte Carlo (MCMC) algorithms, for example, hybrid Monte Carlo[9], which is based on Hamiltonian dynamics.

The dependence of the posterior on the parameters is frequently approximated as a Gaussian distribution in the neighborhood of the maximum a posteriori (MAP) solution, $\hat{\mathbf{a}}$, which minimizes $\varphi(\mathbf{a})$. Thus,

$$\varphi(\mathbf{a}) = \frac{1}{2} (\mathbf{a} - \hat{\mathbf{a}})^T \mathbf{K} (\mathbf{a} - \hat{\mathbf{a}}) + \varphi_0, \quad (1)$$

where \mathbf{K} is the curvature matrix for $\varphi(\mathbf{a})$ and $\varphi_0 = \varphi(\hat{\mathbf{a}})$. The inverse of \mathbf{K} is the covariance matrix, defined as $\mathbf{C} = \text{cov}(\mathbf{a}) = \langle (\mathbf{a} - \hat{\mathbf{a}}) (\mathbf{a} - \hat{\mathbf{a}})^T \rangle$.

EXTERNAL FORCE

In the physics analogy, an external force applied to a physical system in equilibrium will distort it. The displacement is determined by the curvature (or stiffness) matrix describing the potential around $\hat{\mathbf{a}}$. In the inference problem, the idea is to add to $\varphi(\mathbf{a})$ a potential that is linear in \mathbf{a} and find the new minimizer \mathbf{a}' . Equation (1) becomes

$$\varphi'(\mathbf{a}) = \frac{1}{2} (\mathbf{a} - \hat{\mathbf{a}})^T \mathbf{K} (\mathbf{a} - \hat{\mathbf{a}}) - \mathbf{f}^T \mathbf{a} + \varphi_0, \quad (2)$$

where \mathbf{f} is analogous to a force acting on \mathbf{a} . Setting the gradient of $\varphi'(\mathbf{a})$ equal to zero, we obtain

$$\delta \mathbf{a} = \mathbf{a}' - \hat{\mathbf{a}} = \mathbf{K}^{-1} \mathbf{f} = \mathbf{C} \mathbf{f}, \quad (3)$$

since the inverse of \mathbf{K} is \mathbf{C} . This simple relation suggests that one can determine specific elements of \mathbf{C} by selecting the appropriate force vector \mathbf{f} and seeing how the MAP parameters change with reoptimization. If \mathbf{f} has only one nonzero component, f_j , then Eq. (3) becomes $\delta a_j = \sigma_{a_j}^2 f_j$.

Inserting Eq. (3) into Eq. (1), the functional dependence of the change in the posterior is

$$\delta \varphi = \frac{1}{2} (\mathbf{C} \mathbf{f})^T \mathbf{K} (\mathbf{C} \mathbf{f}) = \frac{1}{2} \mathbf{f}^T \mathbf{C} \mathbf{f}, \quad (4)$$

because $\mathbf{K}^{-1} = \mathbf{C}$ and \mathbf{C} is symmetric. Thus, the change in the posterior is quadratically related to the magnitude of the applied force.

Derived quantities

The above technique may be used to estimate the uncertainty in quantities that are derived from the fitted parameters. Suppose the quantity of interest z is a function of the parameters \mathbf{a} . To first order, perturbations in $z(\mathbf{a})$ are given by

$$\delta z = \mathbf{s}_z^T \delta \mathbf{a}, \quad (5)$$

where \mathbf{s}_z is the sensitivity vector of z with respect to \mathbf{a} , $s_i = \frac{\partial z}{\partial a_i}$. The variance in z is

$$\text{var}(z) = \langle |\delta z|^2 \rangle = \langle \mathbf{s}_z^T \delta \mathbf{a} \delta \mathbf{a}^T \mathbf{s}_z \rangle = \mathbf{s}_z^T \mathbf{C} \mathbf{s}_z. \quad (6)$$

Thus, the appropriate force on \mathbf{a} to probe z is $\mathbf{f}_z = k \mathbf{s}_z$, where k is a scaling parameter to adjust the magnitude of the force. Then, $\delta z = \mathbf{C}_z \mathbf{f}_z = \sigma_z^2 k$, which has the same form as Eq. (3). Therefore,

$$\sigma_z^2 = \frac{\delta z}{k}, \quad (7)$$

which can be used to estimate σ_z from the δz produced by the applied force.

From Eq. (4), the dependence of the posterior on \mathbf{f} is

$$\delta \varphi = \frac{1}{2} k^2 \sigma_z^2, \quad (8)$$

or

$$\sigma_z = \frac{\delta z}{\sqrt{2\delta \varphi}}. \quad (9)$$

This relation provides another way to estimate the standard error in z . It is perhaps more reliable than Eq. (7) because it doesn't explicitly involve the magnitude of the force.

EXAMPLES

Fitting a straight line

The first example is very simple; fitting a straight line to measurements in one-dimension. The model for the measurements is $y = a + bx$, where a is the intercept of the line with the y axis and b is the slope of the line. The parameter vector \mathbf{a} consists of the parameters a and b . Figure 1a shows 10 data points obtained for $a = 0.5$ and $b = 0.5$, with additive fluctuations in y produced by drawing random numbers from a Gaussian distribution with zero mean and a standard deviation of 0.2.

Consider now the analysis of these data. Assuming the uncertainties in the measurements are independent and Gaussian distributed, as well as a flat prior, φ for this problem is

$$\varphi(\mathbf{a}) = \frac{1}{2} \chi^2 = \frac{1}{2} \sum_i \frac{[y_i - y(x_i; \mathbf{a})]^2}{\sigma_i^2}, \quad (10)$$

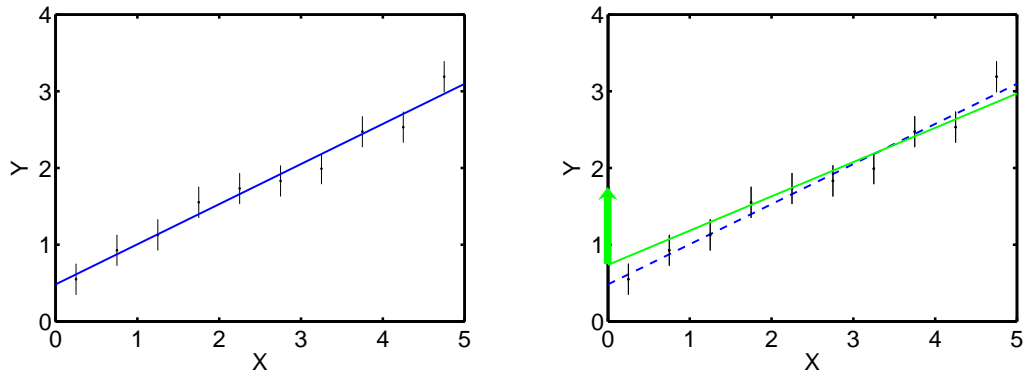


FIGURE 1. (a, left) Plot of 10 data points with their standard error bars, and the straight line that minimizes φ . (b, right) Applying an upward force to the line at $x = 0$ and reoptimizing φ' , lifts the line there. However, the data pull the left side of the line down, resulting in a negative correlation between the intercept, a , and the slope, b .

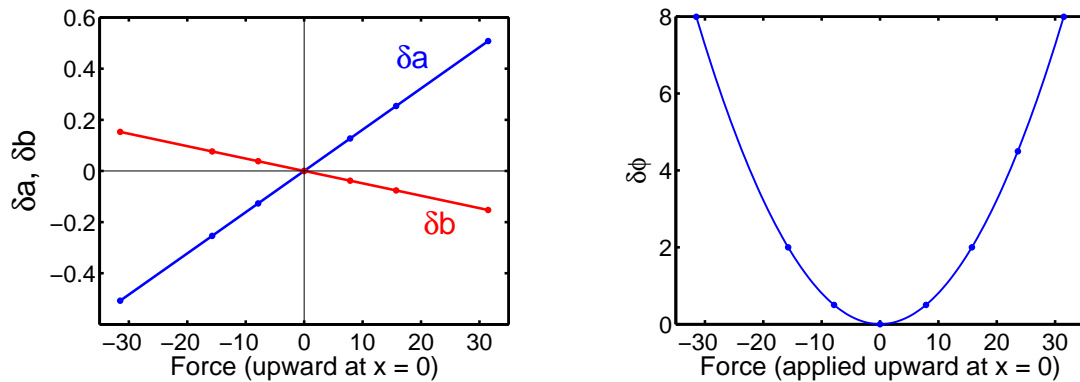


FIGURE 2. (a, left) Plot of the displacements in the two parameters of the straight line in response to an upward force applied to the line at $x = 0$. (b, right) Plot of the change in φ as a function of the magnitude of the applied force. As explained in the text, the functional dependences of either of these plots may be used to quantitatively estimate properties of the covariance matrix.

where y_i is the measurement at the position x_i , and σ_i is its standard error. The line through the data in Fig. 1a represents the MAP solution, that is, it minimizes φ .

Suppose that we apply an upward force to the line at $x = 0$. This force is applied only to the parameter a . The new position of the line obtained by minimizing Eq. (2) is shown in Fig. 1b. The intercept a is increased, while the slope of the line b is decreased to maintain a good fit to the data. This observed anticorrelation between a and b is a direct indication of the correlation between the uncertainties in these variables expressed by the covariance matrix.

Quantitative estimates of elements of \mathbf{C} may be obtained from the plot in Fig. 2a, which shows that the changes in a and b from their original values are proportional to the applied vertical force. The slope of δa relative to f , by Eq. (7) is $C_{aa} = \sigma_a^2 = (0.127)^2$. The slope of δb relative to f is $C_{ab} = -4.84 \times 10^{-3}$. The diagonal term C_{bb} is not

determined because the force chosen does not directly probe that element of \mathbf{C} .

These results may be checked through a conventional least-squares fitting analysis[10]. When the number of parameters is not too large and the function calculations are quick, finite differences may be used to evaluate the Jacobian matrix (derivatives of all outputs with respect to all parameters). The curvature matrix \mathbf{K} in Eq. (1) can be approximated as the outer product of the Jacobian with itself. The inverse of \mathbf{K} is \mathbf{C} .

The conventional analysis confirms that results quoted above.

Spectral fitting

A more complicated example consists of fitting the data shown in Fig. 3. These data are obtained by simulation, assuming the spectrum consists of a single Gaussian peak added to a quadratic background. There are six parameters: the position, amplitude, and rms width of the Gaussian, x_0 , a , and w , respectively, and three parameters to describe the quadratic background. For this example, $x_0 = 3$, $a = 2$, and $w = 0.2$. Random noise is added to the data, assuming $\sigma_y = 0.2$. Figure 3 shows the spectrum obtained by minimizing $\varphi(\mathbf{a}) = \frac{1}{2} \chi^2$ with respect to the six parameters.

Let us assume that we are principally interested in the area under the Gaussian peak. The area is proportional to the product of its amplitude and width: $A = \sqrt{2\pi} a w$. Following the discussion in the section on Derived Quantities, the force to apply to a and w should be proportional to the derivatives of A with respect to these parameters: $\frac{\partial A}{\partial a} = \sqrt{2\pi} w$ and $\frac{\partial A}{\partial w} = \sqrt{2\pi} a$. Examples of the result of applying large positive and negative forces to the area are shown in Fig. 4. Figure 5 shows the results of varying the magnitude of the applied force. For small values of the force, the change in A depends approximately linearly on the force. However, for this nonlinear model, the linear dependence is expected to fail at some point.

From Eq. (9), the quadratic dependence of $\delta\varphi$ on δA should be $\delta\varphi = \frac{1}{2} \left[\frac{\delta A}{\sigma_A} \right]^2$. For the smallest forces applied, this yields the estimates $\sigma_A = 0.098$, for a negative force and 0.104, for a positive force.

The results of a conventional least-squares analysis are $\chi_{\min}^2 = 34.32$ with $p = 0.852$; $a = 1.948$, $\sigma_a = 0.149$, $w = 0.1759$, $\sigma_w = 0.0165$, and a correlation coefficient $r_{aw} = -0.427$. From these, the area is $A = 0.859$, and its standard error is $\sigma_A = \sqrt{2\pi} [w^2 \sigma_a^2 + a^2 \sigma_w^2 - r_{aw} a w \sigma_a \sigma_w]^{1/2} = 0.093$, in reasonable agreement with the above result, considering the slightly nonquadratic behavior of the curve.

Tomographic reconstruction

The foregoing examples were simple enough to be handled by standard least-squares fitting routines. However, as the number of variables increases, let us say beyond several hundred, and as the cost of function evaluation increases, standard fitting codes fail. The general approach that has developed avoids full-blown matrices; the MAP estimates are found by numerical optimization of the calculated φ . The difficult part is estimating

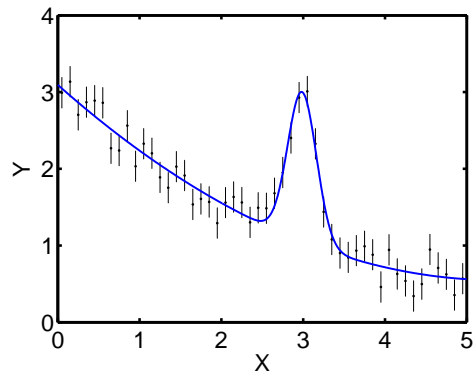


FIGURE 3. Plot of 50 data points with their standard error bars representing a simple spectrum, and the optimized curve corresponding to a Gaussian peak on a quadratic background.

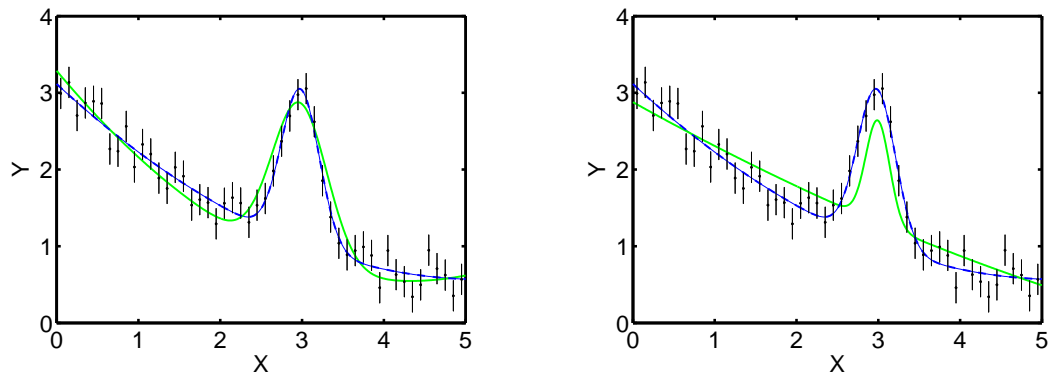


FIGURE 4. (a, left) Plot of the reoptimized curve under the influence of an external force applied to the parameters to increase the area under the Gaussian peak. (b, right) Similar to a, but the external force is applied to the parameters to decrease the area under the peak. The dashed lines are the unperturbed curve.

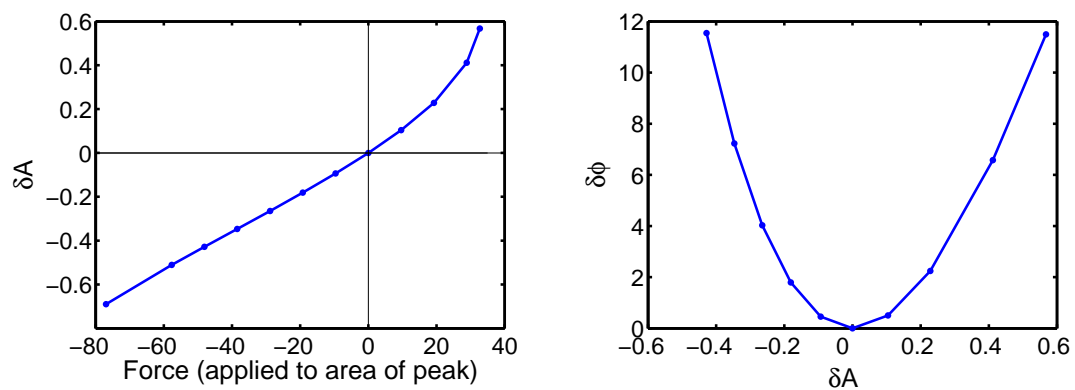


FIGURE 5. (a, left) Plot of the change in the area under the peak A caused by a force applied to A . (b, right) Plot of the change in φ as a function of the change in the area produced by the applied force.

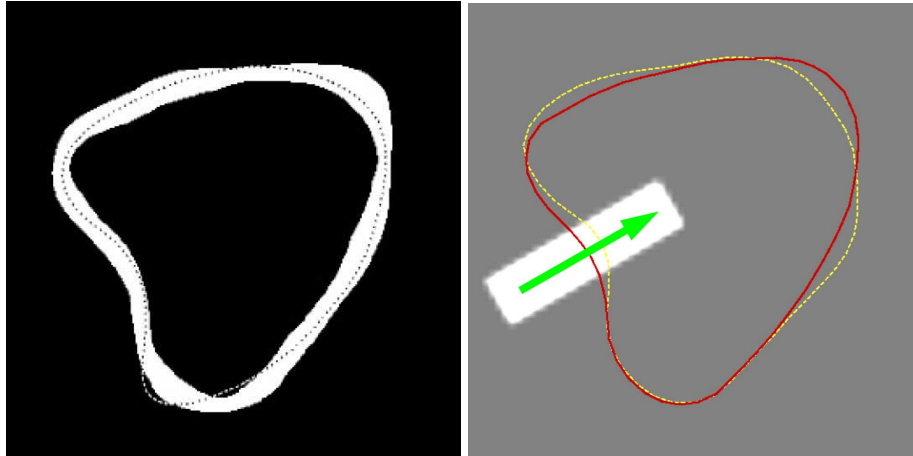


FIGURE 6. (a, left) Display of the 95% Bayesian credible interval for the boundary of an object reconstructed from two projections obtained using MCMC and the Bayes Inference Engine (from Ref. [7]). The white region corresponds to roughly a plus or minus two-standard-deviation uncertainty envelop for the edge of the object. The boundary of the original object in this test problem is shown as a dashed line. (b, right) The dashed (yellow) line shows the effect of pushing inward on the MAP-estimated boundary, shown as the solid (red) line. The displacement of the boundary corresponds to the covariance matrix for the boundary location, relative to the position of the applied force, indicated by the white rectangle.

the uncertainties in the parameters. For that purpose, Markov Chain Monte Carlo[11] (MCMC) is often employed in Bayesian calculations. Indeed, it is very adaptable and handles virtually every situation. However, MCMC tends to be relatively inefficient and time consuming under the conditions described.

The Bayes Inference Engine[12] (BIE) was developed at the Los Alamos National Laboratory to provide an effective means for tomographically reconstructing objects from radiographs. An example of the use of the BIE was presented in Ref. [7], which demonstrated the reconstruction of an object from just two orthogonal projections. The object was modeled in terms of a deformable boundary with a known constant interior density. The BIE optimized the simple boundary to match the two noisy projections by means of adjoint differentiation, which efficiently provides the derivatives of the optimization function with respect to the 50 vertices of the polygonal boundary in the same amount of time as one forward calculation. Thus, a gradient-based optimization algorithm, BFGS[13], can be used to accelerate convergence. In the above article, MCMC, in the form of the Metropolis algorithm[3], was used to assess the uncertainties in the edge location. The end result of that example was the 95% Bayesian credible interval for the boundary of the object shown in Fig. 6a.

In this situation, the concept of probing the covariance matrix is ideally suited to determining the uncertainty in the edge location of the reconstruction at a particular position. Figure 6b shows the result of applying pressure (force over the finite width of the white rectangle) to the boundary. The solid (red) line represents the MAP estimated boundary. The dashed (yellow) line shows the new position of the boundary, after minimizing the augmented posterior (2).

The deflection of the boundary over the width of the rectangle may be used to

quantitatively estimate the standard error in the estimated edge location at that spot. Furthermore, the deflections elsewhere are proportional to the covariance between those edge locations and the probed position. The main effects in these observed correlations are easy to understand, given that the measurements on which the reconstruction are based consist in horizontal and vertical projections of the object's density distribution. The edge locations opposite to the pressure point, horizontally and vertically, move outwards to maintain the projection values. The inward movement of the boundary in the upper right-hand position of the object is similarly in response to the latter two motions.

SUMMARY

I have presented a novel method for numerically estimating elements of the covariance matrix. The method relies on optimization of the minus-log-posterior, and so replaces standard stochastic methods with a deterministic one.

The method consists of the following steps:

1. Find the model parameters $\hat{\mathbf{a}}$ that minimize φ (minus-log-posterior).
2. Decide on the quantity of interest z .
3. Calculate the sensivity of z with respect to \mathbf{a} : $s_z = \frac{\partial z}{\partial \mathbf{a}}$.
4. Find the parameters that minimize $\varphi' = \varphi - k \mathbf{s}_z^T \mathbf{a}$. The factor k should be approximately σ_z^{-1} . If $\delta\varphi$, is much bigger than 0.5, reduce k and try again.
5. Estimate the standard error in z with either: $\sigma_z^2 = \frac{\delta z}{k}$ or $\sigma_z = \frac{\delta z}{\sqrt{2\delta\varphi}}$.

Furthermore, the covariance between z and other quantities may be estimated using Eq. (3).

The described method may be most useful when: a) one's interest is in the uncertainty in one or a few parameters or derived quantities, out of many parameters; b) the full covariance matrix is not known (nor desired); c) the posterior can be well approximated by Gaussian distribution in parameters; and d) minimization of φ and φ' can be done efficiently. The latter condition seems to require that the gradient calculation can be done efficiently, for example, through adjoint differentiation of the forward simulation code[14]. Some potential uses of the method include estimation of the signal-to-noise ratio in a region of a tomographic reconstruction[15, 16] and estimating the uncertainty in the scalar output of a simulation code, for example, the criticality of an assembly of fissile material calculated with a neutron-transport code[17].

The method may also be useful for exploring and quantifying non-Gaussian posterior distributions, including situations with inequality constraints. For example, non-negativity constraints in an inverse problem may result in some parameters being pushed to the limit. The gradient at the MAP solution may not be zero because of the constraint. Using a force to probe the posterior can quantify the strength of the constraint. Likelihoods with long-tails are often used to handle outliers[18]. Such likelihoods can lead to nonGaussian posterior distributions. The present method may be used to explore such nonGaussian distributions, even though there may be no interpretation in terms of covariance. The method might be useful to explore probabilistic correlations in self-optimizing natural systems, such as populations, bacteria, and traffic.

ACKNOWLEDGMENTS

Greg Cunningham played a major role in developing the original technique and implementing it in the Bayes Inference Engine. I would like to thank Rainer Fischer and Paul Goggans for useful discussions, and Richard Silver and Lawrence Pratt for answering statistical-physics questions. This work was done under U.S. DOE Contract DE-AC52-06NA25396.

REFERENCES

1. D. S. Sivia and J. Skilling, *Data Analysis - A Bayesian Tutorial: Second Edition*, Clarendon, Oxford, 2006.
2. T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*, Springer, New York, 2003.
3. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, 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. K. M. Hanson and G. S. Cunningham, "The hard truth," in *Maximum Entropy and Bayesian Methods*, J. Skilling and S. Sibisi, eds., pp. 157–164, Kluwer Academic, Dordrecht, 1996.
6. G. S. Cunningham and K. M. Hanson, "Uncertainty estimation for Bayesian reconstructions from low-count SPECT data," in *Conf. Rec. IEEE Nucl. Sci. Symp. and Med. Imaging Conf.*, IEEE, Piscataway, 1996.
7. K. M. Hanson, G. S. Cunningham, and R. J. McKee, "Uncertainty assessment for reconstructions based on deformable models," *Int. J. Imaging Systems and Technology* **8**, pp. 506–512, 1997.
8. L. E. Reich, *A Modern Course in Statistical Physics - Second Edition*, Wiley, New York, 1998.
9. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," *Phys. Lett. B* **195**, pp. 216–222, 1987.
10. P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, Boston, 1992.
11. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, London, 1995.
12. K. M. Hanson and G. S. Cunningham, "Operation of the Bayes Inference Engine," in *Maximum Entropy and Bayesian Methods*, W. von der Linden et al., ed., pp. 309–318, Kluwer Academic, Dordrecht, 1999.
13. P. E. Gill, W. Murray, and F. H. Wright, *Practical Optimization*, Academic, New York, 1981.
14. A. Griewank, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, SIAM, Philadelphia, 2000.
15. J. Qi and R. H. Huesman, "Theoretical study of penalized-likelihood image reconstruction for region of interest quantification," *IEEE Trans. Med. Imaging* **25**, pp. 640–648, 2006.
16. K. M. Hanson and K. J. Myers, "Rayleigh task performance as a method to evaluate image reconstruction algorithms," in *Maximum Entropy and Bayesian Methods*, W. T. Grandy and L. H. Schick, eds., pp. 303–312, Kluwer Academic, Dordrecht, 1991.
17. T. Kawano, K. M. Hanson, S. C. Frankle, P. Talou, M. B. Chadwick, and R. C. Little, "Uncertainty quantification for applications of ^{239}Pu fission cross sections using a Monte Carlo technique," *Nucl. Sci. Engr.* **153**, pp. 11–7, 2006.
18. K. M. Hanson, "Bayesian analysis of inconsistent measurements of neutron cross sections," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, K. H. Knuth et al., ed., *AIP Conf. Proc.* **803**, pp. 431–439, AIP, Melville, 2005.