THE HARD TRUTH

Kenneth M. Hanson and Gregory S. Cunningham*
Los Alamos National Laboratory, MS P940
Los Alamos, New Mexico 87545 USA
kmh@lanl.gov cunning@lanl.gov

ABSTRACT. Bayesian methodology provides the means to combine prior knowledge about competing models of reality and available data to draw inferences about the validity of those models. The posterior quantifies the degree of certainty one has about those models. We propose a method to determine the uncertainty in a specific feature of a Bayesian solution. Our approach is based on an analogy between the negative logarithm of the posterior and a physical potential. This analogy leads to the interpretation of gradient of this potential as a force that acts on the model. As model parameters are perturbed from their maximum a posteriori (MAP) values, the strength of the restoring force that drives them back to the MAP solution is directly related to the uncertainty in those parameter estimates. The correlations between the uncertainties of parameter estimates can be elucidated.

1. Introduction

Bayesian analysis provides the foundation for a rich environment in which to explore inferences about models from both data and prior knowledge through the posterior probability. In an attempt to reduce an analysis problem to a manageable size, the usual approach is to present a single instantiation of the object model as "the answer", typically that which maximizes the posterior (the MAP solution). However, because of uncertainties in the measurements and/or because of a lack of sufficient data to define an unambiguous answer (in the absence of regularizing priors) [1], there is no unique answer to many real analysis problems. Rather, innumerable solutions are possible. Of course, some solutions are more probable than others. The beauty of the Bayesian approach is that it provides the probability of every possible solution, which, in a sense, ranks various solutions. The estimation of the uncertainty or reliability of the answer remains a pressing issue, particularly when the number of parameters in the model is large. Although there is a mathematically correct way to specify the covariance in the parameters, including the correlation between the uncertainties in any two parameters, it does not provide much insight.

One appealing way to get a feeling for the uncertainty in a Bayesian solution is to display a sequence of distinct solutions drawn from the posterior probability distribution. This approach was suggested by Skilling et al. [2], who produced a video display of a random walk through the posterior distribution. However, the calculational method used in that work was based on a Gaussian approximation of the posterior probability distribution in the neighborhood of the MAP solution. Later Skilling made some progress in dealing with non-Gaussian distributions [3]. While the probabilistic display of Skilling et al. provides

*Supported by the United States Department of Energy under contract number W-7405-ENG-36.

a general impression of the overall degree of possible variation in the solution, we desire a means to probe the uncertainty in the solution in a more directed manner.

We propose a technique to test hypotheses regarding perturbations of the MAP solution in a fashion that allows one to ask questions of particular interest. The approach we suggest makes use of an analogy between the negative logarithm of the posterior and a physical potential. The uncertainty of a particular change of the MAP solution is revealed in a tactile way as a force that tends to pull the solution back toward the MAP solution. Correlations between the perturbed set of parameters and the remaining parameters in the model are also brought to light. This innovative Bayesian tool is tangibly demonstrated within the context of geometrically-defined object models used for tomographic reconstruction from very limited projection data.

2. Traditional approach to uncertainty

Bayesian analysis revolves around the posterior probability of a model, where the model parameters are represented by the vector $\mathbf{a}$. The posterior $p(\mathbf{a}|\mathbf{d})$ incorporates data through the likelihood $p(\mathbf{d}|\mathbf{a})$, i.e. the probability of the observed data given the parameters, and prior information through a prior probability on the parameters $p(\mathbf{a})$. Bayes's law gives the posterior as $p(\mathbf{a}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{a})p(\mathbf{a})$. The most typical use of Bayesian analysis is to find the parameter values that maximize the posterior, called the MAP solution.

It is convenient to deal with the negative logarithm of the posterior, $\varphi = -\log\{p(\mathbf{a}|\mathbf{d})\}$. The MAP estimate of the parameters $\hat{\mathbf{a}}$ is found by minimizing φ, the condition for which is $\frac{\partial \varphi}{\partial a_i} = 0$ for all parameters a_i, providing there are no constraints on the parameters themselves. In the traditional approach to the estimation of uncertainty [4], which we only summarize here, the variances in the parameters are derived from the curvature matrix of φ, calculated at $\hat{\mathbf{a}}$,

$$K_{ij} = \frac{\partial^2 \varphi}{\partial a_i \partial a_j} \quad . \tag{1}$$

Since this matrix is evaluated at the minimum of φ, it must be positive semi-definite, i.e. $(\Delta\mathbf{a})^{\mathrm{T}}\mathbf{K}\,\Delta\mathbf{a} \geq 0$ for any $\Delta\mathbf{a}$. In the Gaussian approximation to the posterior, the error matrix $\mathbf{E}$, which gives the covariances between all the parameters, $[\mathbf{E}]_{ij} = \langle (a_i - \hat{a}_i)(a_j - \hat{a}_j) \rangle$, where the brackets indicate an ensemble average, is the inverse of the curvature matrix,

$$\mathbf{E} = \mathbf{K}^{-1} \quad . \tag{2}$$

Although this result is mathematically rigorous, it only provides the second moment of the parameter uncertainties and their correlations. It also suffers from not being very illuminating in terms of its consequences for the parametric model, particularly in terms of the correlations in the uncertainties of various parameters. Furthermore, for the 10^5–10^6 pixel amplitudes that are typically needed to describe a 2D image, the full error matrix contains 10^{10}–10^{12} elements, which can neither be practically calculated nor stored. We propose another approach to provide a more tangible indication of the degree of uncertainty in the inferred model as well as the ability to directly probe the uncertainty of specific features of the model.

3. Bayesian mechanics

If one draws an analogy between φ and a physical potential, then the gradient of φ is analogous to a force, just as in physics. The force $f_i = -\frac{\partial \varphi}{\partial a_i}$ is roughly in the direction of the local minimum of φ, under suitable assumptions concerning the smoothness of the dependence of φ on the parameters. The condition for the MAP solution, $\nabla_{\mathbf{a}} \varphi = 0$, can be interpreted as stating that at the MAP operating point the forces on all the variables in the problem balance: the net force on each variable is zero. Further, when the variable a_i is perturbed slightly from the MAP solution, the force f_i pulls a_i back towards the MAP solution. The phrase "force of the data" takes on real meaning in this context.

A quadratic approximation to φ in the neighborhood of the MAP solution implies a linear force law, i.e. the restoring force is proportional to the displacement from equilibrium, as in a simple spring. In this quadratic approximation the curvature of φ is proportional to the covariance of the MAP estimate. A high curvature is analogous to a stiff spring and therefore represents a "rigid", reliable solution.

An interesting aspect of this interpretation is the possibility of decomposing the forces acting on the MAP solution into their various components. For example, the force derived from all data (through the likelihood), or even a selected set of data, may be compared to the force derived from the prior. In this way it is possible to examine the influence of the priors on the solution as well as determine which data have the largest effect on a particular feature of the solution.

We note that the notion of applying forces to model parameters in the preceding discussion must ultimately be stated in terms of pressures, that is, forces applied over regions, acting on physically meaningful quantities. The first reason is that the physical world, which we usually model, exists as a continuum: the physical quantities of interest are typically densities, which are a function of continuous spatial or temporal coordinates. Thus meaningful questions about reality should really be stated in terms averages over regions, not as point values. Secondly, physically feasible measurements can only probe physical quantities over finite-sized regions. Point sampling is fundamentally impossible. As an example, a radiographic measurement in which the attenuation of an x-ray beam is measured is always subject to the effects of a blurring process that arises from a finite spot size for the source of x rays and the finite resolution of the x-ray detector. Thus the measured attenuation is necessarily an average over a cylinder in space. In truth, radiographic measurements can not provide line integrals of an attenuation coefficient through an object, as is often assumed as an approximation to the real process. Put succinctly, all physical measurements have limited spatial or temporal resolution that render as meaningless questions about what happens in an infinitesimally small region. As a result, uncertainties in an estimated physical quantity can only be addressed in terms of the average of that quantity over a finite region. As the concepts of Bayesian analysis mature, we will learn to deal only with physical quantities that are functions of continuous independent variables and we will avoid referencing directly the underlying discrete parameters of the models.

One needs to be aware that any finite representation, which we are forced to use in computer models, has a limited resolution. Thus when one explores the model at a scale finer than the inherent resolution of the model, the model can only respond by interpolation of the underlying discrete model [5]. One can only meaningfully explore the model at resolutions coarser than this.

4. Perturbation from Equilibrium

We propose to exploit the above physical analogy to facilitate the exploration of the uncertainty in a MAP solution. For the present we will assume that in the neighborhood of the MAP point $\hat{\mathbf{a}}$ φ is well approximated by a quadratic expansion:

$$\varphi = \tfrac{1}{2}(\Delta\mathbf{a})^{\mathrm{T}}\mathbf{K}\,\Delta\mathbf{a} + \varphi_0 \;, \tag{3}$$

where $\Delta\mathbf{a} = \mathbf{a} - \hat{\mathbf{a}}$ is the displacement from the MAP point and $\varphi_0 = \varphi(\hat{\mathbf{a}})$. Suppose that we start from $\hat{\mathbf{a}}$ and displace the parameter values by a small amount $\Delta\mathbf{a}$. Then the gradient of φ, $-\nabla_{\mathbf{a}}\varphi$, represents a force that pulls the parameters back toward the MAP point. The units of the force are those of the reciprocal of the parameter. The curvature, and hence the reciprocal of the variance, in the direction of $\Delta\mathbf{a}$ is given by the ratio of $|\nabla_{\mathbf{a}}\varphi|$, evaluated at $\hat{\mathbf{a}} + \Delta\mathbf{a}$, to $|\Delta\mathbf{a}|$, for vanishingly small displacements.

As an alternative to directly displacing parameters, their perturbation may be achieved by applying an external force to the parameters. Suppose that one pulls on the parameters with a force $\mathbf{f}$. Note that this force can act on just one parameter or on many. From the physical analogy, it is easy to write down the new potential;

$$\varphi = \tfrac{1}{2}(\Delta\mathbf{a})^{\mathrm{T}}\mathbf{K}\,\Delta\mathbf{a} - \Delta\mathbf{a}^{\mathrm{T}}\mathbf{f} + \varphi_0 \;. \tag{4}$$

The new minimum of φ occurs when

$$\nabla_{\mathbf{a}}\varphi = 0 = \mathbf{K}\Delta\mathbf{a} - \mathbf{f} \;. \tag{5}$$

Solving for the displacement in $\mathbf{a}$ and using Eq. (2),

$$\Delta\mathbf{a} = \mathbf{K}^{-1}\mathbf{f} = \mathbf{E}\mathbf{f} \;. \tag{6}$$

If the curvature matrix $\mathbf{K}$, and hence the covariance matrix $\mathbf{E}$, is not diagonal, the resulting displacement is not in the direction of the applied force. This phenomenon demonstrates the correlations between the uncertainties in all the parameters. The component of $\Delta\mathbf{a}$ in the direction of the applied force divided by the magnitude of the force, i.e. $\Delta\mathbf{a}^{\mathrm{T}}\mathbf{f}/|\mathbf{f}|^2$, is the effective variance in the parameters in that direction.

Although we assumed that φ is quadratic above, this approach can be useful even when it is nonquadratic. While it may not be feasible to express the results analytically, we obtain a feeling for the uncertainty in $\Delta\mathbf{a}$ and the correlations between $\Delta\mathbf{a}$ and the other parameters. Any constraints on the parameters can be explicitly seen. For nonquadratic φ the plot of the value of φ versus the applied force provides the means to visualize the uncertainty in $\Delta\mathbf{a}$.

5. Use with Deformable Geometric Model

The above approach takes on a poignant interpretation when the reconstructed object is defined in terms of its geometric shape. The prior on the geometry is defined in terms of the default shape together with a prescription of how to assess the probability of other possible shapes. The latter is simply done by using a Gibbs form for the probability given as $\exp(-\beta W)$, where W is the deformation energy, i.e. the energy required to deform the

Figure 1: An example of how a polygon (solid line) can be distorted by either pushing node A inward (dashed line) or outward (dotted line), assuming that the measurements consist of two orthogonal projections. Note the effect on the overall shape of the object, which indicates the correlations between the polygon vertices.

geometry from the default shape into a new shape [6–10]. The parameter β regulates the strength of the prior on the geometry.

Figure 1 shows a polygon defined in terms of its 20 vertices or nodes. Thus there are 40 parameters in this model corresponding to the two coordinates needed to specify each vertex of the polygon. We assume that two sets of parallel projections, one vertical and one horizontal, are available and that they are subject to a very small amount of measurement noise. For simplicity we ignore the prior on the deformation described above. Starting from the known original polygon, a force is applied to the leftmost node (node A), pulling it outward. The plot of the applied force and the resulting horizontal displacement of the node is shown in Fig. 2. For positive forces node A moves outward steadily up to a breakpoint (at a displacement of 0.18), which we call point B. The dotted-line figure in Fig. 1 shows the configuration of the polygon at that point. We note that the act of displacing node A outward contradicts the vertical projections, which indicate that there is probably no material to the left of the original position of the node. Beyond point B the slope of the curve decreases substantially, principally because new configurations of the polygon are possible, which can reduce the excessive projection values to the left of the original position of node A. As an aside, the optimization procedure employed is based on a steepest descent method. We use the technique of adjoint differentiation to efficiently calculate the required derivatives with respect to the parameters [11].

Figure 2: Plot of the force applied to node A of the polygon in Fig. 1 versus the resulting displacement of that node. The nonlinear nature of the force-displacement law for this problem is dramatically demonstrated. The configurations shown in Fig. 1 are at the two breakpoints in the curve: the dashed line corresponds to a force of -0.006 (inward) at point C and the dotted line to a force of 0.080 (outward) at B.

Applying the force inward (negative force values) results in quite a different behavior. With a small inward push, the displacement reaches a breakpoint, point C in Fig. 2. The configuration of the polygon at this point is shown Fig. 1 as the dashed figure. Node A has just reached the line connecting its neighbors, one of which has moved outward to take its place in supplying the proper vertical projection. Pushing harder only makes node A slide down that line, which requires only a little force to achieve a large displacement. The position of node A is not well determined in this region. We notice that the shape of the object does not change during this process. The results for this situation are correct, but may not be what one has in mind when specifying the force. It seems desirable to avoid applying the force directly to the parameters, in this case, to the position of the nodes of the polygon. The force should instead be applied to the object and its effect translated to the parameters. Also we observe that the only reason point C is not closer to the origin is that the coarseness of our polygon object model limits the flexibility of the object to respond. With many more degrees of freedom, we would expect neighboring sections of the object boundary to move out to take the place of node A in response to a slight inward force.

The correlation between the uncertainty in the position of node A and the positions of the other nodes in the polygon is demonstrated in Fig. 1. We observe that the nodes on the right side of the polygon move to maintain the measured horizontal projection. Of course, the constraints of the vertical projection also figure into the problem, making the overall movement of the sides of the polygon rather complex. This approach nicely handles the complex interaction between all the constraints arising from measurements and prior knowledge.

For an object modeled in terms of its geometry, poor reliability of the MAP estimate means that the object is soft or squishy, pliable. Good reliability of the estimate means that the object is firm. Therefore, "truth" is hard or rigid.

6. Discussion

In the future it may be possible to use the tools of virtual reality, coupled to turbocomputation, to explore the reliability of a Bayesian solution of complex problems through direct manipulation of the computer model. Force feedback will permit one to actually "feel" the stiffness of a model. Higher dimensional correlations might be "felt" through one's various senses.

To reiterate the comments made in Sect. 3, we suggest that queries regarding physical quantities should be made in terms of averages over regions rather than in terms of their values at points. Furthermore, the uncertainties of individual parameters that, as a collection, are meant to describe a physical quantity as a function of continuous coordinates, may have little meaning. In regard to an image represented as a grid of pixels, the question "what is the rms error in a pixel value?," is impossible to answer without a clear understanding of what a pixel value represents, e.g. the average value over the area of the pixel. More meaningful questions can be made for areas larger than that of a single pixel. Furthermore, the correlations between the average value within a region and the rest of the image must be considered. Consequently, our language must change. Instead of applying forces to probe the reliability of individual parameters that are used to describe an object, we should speak of applying pressures over regions of the object. And it must be understood that when we ask about regions whose size is on the order of, or smaller, than the resolution of the discrete model of the object, we will only learn about the interpolation properties of the model.

The approach to reliability testing described above is very general and can be used in virtually any other kind of Bayesian analysis. Examples of other contexts are as follows:

Spectral estimation: In typical spectral analysis a scalar variable quantity is estimated for different discrete frequency values. Normally a single spectrum is estimated. Skilling et al. [3] probed the variability possible in the answer through their probabilistic display technique. That display gives one a true feeling for the range of answers possible for a given set of input data. With our technique, one can ask direct questions about the power over a specific range of frequencies. The mode of interaction with the spectrum might be thought of as pushing down or pulling up on a region of the spectrum. In a virtual reality setting, we can imagine that the analyst would be able to use his fingers to press upward or downward on various portions of the spectrum. The resistance to this attempted action, fed back to the users fingers as a force, would indicate the degree of uncertainty in the solution.

Image reconstruction: The basic problem is to estimate the amplitudes in image pixels from data, each of which is a combination of many pixels, as in tomographic reconstruction from projections (strip integrals) through the image, or deconvolution of blurred images. Interaction with the image can be provided by allowing one to push or pull on the amplitudes in an area of interest. The concepts behind this technique can be used to make binary decisions, for example, to decide whether an object is present or not, or to decide between two different signals [12].

References

[1] K. M. Hanson and G. W. Wecksung. Bayesian approach to limited-angle reconstruction in computed tomography. *J. Opt. Soc. Amer.*, 73:1501–1509, 1983.

[2] J. Skilling, D.R.T. Robinson, and S.F. Gull. Probabilistic displays. In Jr. W.T. Grandy and L.H. Schick, editors, *Maximum Entropy and Bayesian Methods*, pages 365–368. Kluwer Academic, 1991.

[3] J. Skilling. Clouds. presented at the Workshop on Maximum Entropy and Bayesian Methods, July 19-24, 1992, Paris.

[4] P. R. Bevington. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, New York, 1969.

[5] K. M. Hanson and G. W. Wecksung. Local basis-function approach to computed tomography. *Appl. Opt.*, 24:4028–4039, 1985.

[6] R. Szeliski. Probabilistic modeling of surfaces. *Proc. SPIE*, 1570:154–165, 1991.

[7] R. Szeliski and D. Terzopoulos. Physically-based and probabilistic models for computer vision. *Proc. SPIE*, 1570:140–152, 1991.

[8] K. M. Hanson. Reconstruction based on flexible prior models. *Proc. SPIE*, vol. 1652:183–191, 1992.

[9] K. M. Hanson. Flexible prior models in Bayesian image analysis. In A. Mohammad-Djafari and G. Demoment, editors, *Maximum Entropy and Bayesian Methods*, pages 399–406. Kluwer Academic, 1993.

[10] K. M. Hanson. Bayesian reconstruction based on flexible prior models. *J. Opt. Soc. Amer.*, A10:997–1004, 1993.

[11] G. S. Cunningham, K. M. Hanson, G. R. Jennings, Jr., and D. R. Wolf. An object-oriented optimization system. *Proc. IEEE Int. Conf. Image Processing*, III:826–830, 1994.

[12] K. M. Hanson. Making binary decisions based on the posterior probability distribution associated with tomographic reconstrcutions. In C. R. Smith, G. J. Erickson, and P. O. Neudorfer, editors, *Maximum Entropy and Bayesian Methods*, pages 313–326. Kluwer Academic, 1991.