Rezghi, M., and Hosseini, S. M., 2009. A new variant of L-curve for Tikhonov regularization. *Journal of Computational and Applied Mathematics*, **231**, 914–924.
Rojas, M., and Sorensen, D. C., 2002. A trust region approach to the regularization of large scale discrete forms of ill-posed problems. *SIAM Journal of Scientific Computing*, **23**(6), 1843–1861.
Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F., 2006. Sensitivity analysis practices: strategies for model based inferences – review. *Reliability Engineering and Safety*, **91**, 1109–1125.
Scales, J. A., Smith, M. L., and Treitel, S., 2001. *Introductory Geophysical Inverse Theory*. Golden, White River Junction: Smizdat.
Sneider, R., and Trampert, J., 1999. Inverse problems in geophysics. In Wirgin, A. (ed.), *Wavefield Inversion*. New York: Springer, pp. 119–190.
Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia: SIAM.
Tikhanov, A. N., and Arsenin, V. Y., 1977. *Solutions Ill-Posed Problems*. New York: Wiley.
Treitel, S., and Lines, L., 2001. Past, present, and future of geophysical inversion – a new millennium analysis. *Geophysics*, **66**(1), 21–24.
Twomey, S., 1977. *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*. Amsterdam: Elsevier Scientific.
Zhdanov, M. S., 2002. *Geophysical Inverse Theory and Regularization Problems*. New York: Elsevier.

Cross-references

Inverse Theory, Artificial Neural Networks
Inverse Theory, Global Optimization
Inverse Theory, Singular Value Decomposition

INVERSE THEORY, MONTE CARLO METHOD

Malcolm Sambridge[1], Kerry Gallagher[2]
[1]Seismology and Mathematical Geophysics, Research School of Earth Sciences, The Australian National University, Canberra, Australia
[2]UMR 6118- Géosciences Rennes, Gêosciences, Université de Rennes 1, Rennes Cedex, France

Definition

Monte Carlo method. A computational technique making use of random numbers to solve problems that are either probabilistic or deterministic in nature. Named after the famous Casino in Monaco.
Monte Carlo inversion method. A method for sampling a parameter space of variables representing unknowns, governed by probabilistic rules.
Markov chain Monte Carlo (McMC). A probabilistic method for generating vectors or parameter variables whose values follow a prescribed density function.

Introduction

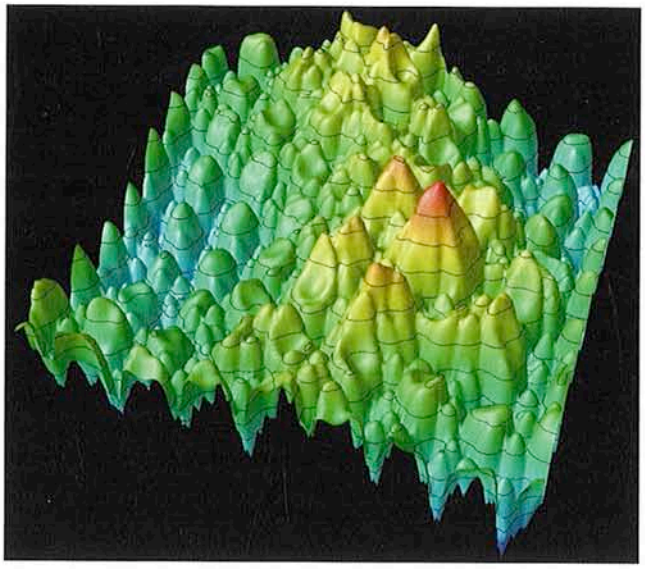
Because geophysical observations are made at (or very near) the Earth's surface, all knowledge of the Earth's interior is based on indirect inference. There always exists an inverse problem where models of physical properties are sought at depth that are only indirectly constrained by the available observations made at the surface. Geophysicists have been dealing with such problems for many years, and in doing so have made substantial contributions to the understanding of inverse problems.

Pioneering work on linear inverse problems in the 1960s arose out of the need to understand how to use new surface observables from seismology to constrain radial variations in geophysical properties at depth within the Earth. Data were few in number and attention was focused on the mathematical structure of the inverse problem and the ways in which reliable information could be recovered. This resulted in a series of important papers beginning with Backus and Gilbert (1967, 1968, 1970). Since that time the geosciences, like many other fields, have moved into a data-rich environment with increasing availability of computational power. Considerable progress has been made over 30 years utilizing the class of linear (typically least squares) parameter estimation algorithms, which are common to many areas of the physical sciences (Aster et al., 2005). In many of the inverse problems encountered the dependence of data on models is nonlinear and this must be taken into account for meaningful solutions. Often this is achieved by performing a local linearization and using *Inverse Theory, Linear*. As the mathematical relationship between data and unknowns becomes complex then linearized methods fail because they depend heavily on having a starting model for the iterative process which must be close enough to the solution for convergence.

Over the last 30 years there has been considerable progress in the solution of highly nonlinear inverse problems involving a limited number of parameters so that thorough exploration can be made of the character of models. Many algorithms have been devised, most of which make use of random numbers to make decisions, that is, in how to generate a set of values of the unknowns whose predictions can be compared to the available data. The original description of Monte Carlo methods by Hammersley and Handscomb (1964) is "the branch of experimental mathematics which is concerned with experiments on random numbers." By this definition all inversion techniques that make use of random numbers are Monte Carlo methods. Many of the *Inverse Theory, Global Optimization* inversion methods fall within the class. The particular approach known as Markov chain Monte Carlo is the primary focus of the present article.

Nonlinearity and multimodal fitness functions

Figure 1 shows a fitness surface from an inverse problem that arises in the analysis of infrasound array data. (Infrasound arrays are used by the United Nations Comprehensive Test Band Treaty Organization to monitor international adherence to the nuclear test ban treaty.) The height of the surface represents the degree of



Inverse Theory, Monte Carlo Method, Figure 1 A multimodal data fit surface arising from the mismatch between two oscillatory fields in the infrasound inversion problem (Kennett et al., 2003).

agreement between two oscillatory fields. There are just two unknowns in this case, which represent tuning parameters in the infrasound array. The object of the exercise is to best tune the array for sensitivity to incoming atmospheric signals, which means finding the point on the surface where the fitness is maximum. We see a curved rim of local maxima from background to foreground, with a broader valley of low fit to the left of the central maximum (red). The physics of the forward problem, that is, calculating the oscillatory field corresponding to a pair of tuning parameters, as well as the nature of the data itself result in a complicated 2-D fitness function.

The set of unknowns that gives the best fit (i.e., smallest misfit) to data corresponds to the global maximum of the multimodal function and to find it one must employ *Inverse Theory, Global Optimization* techniques such as model space search. In this example the global maximum (at the red central peak) was efficiently found with the neighborhood algorithm of Sambridge (1999), which utilizes ideas from the field of computational geometry. Optimization techniques based on local *Inverse Theory, Linear* would only be suitable once a trial solution is found within the vicinity of the global peak (shaded red). Adaptive Monte Carlo-based direct search approaches like genetic algorithms, simulated annealing, and the neighborhood algorithm (see *Inverse Theory, Global Optimization*) are able to solve this (two unknown parameters) problem relatively easily due to their ability to detect the variation of the fit and concentrate sampling where there is most benefit.

This example demonstrates the complexity of inverse problems in cases where the data are highly oscillatory waveforms, a common situation in fields such as acoustics and seismology, where the dimension of the problem is often much higher.

For the 2-D example in Figure 1, the objective is to tune a particular instrument for maximum sensitivity, and it is appropriate to seek a global maximum. More generally in inverse problems the fitness landscape would represent the difference between observations and predictions made by a mathematical model. In this case simply finding the best-fit solution is inadequate. One needs to characterize the uncertainty in the solution, for example, assess how noise in the data lead to errors and trade-offs in the estimated model. Linearized techniques (see *Inverse Theory, Linear*) could be used, but all uncertainty estimates are then based on the assumption of local linearity and do not truly reflect the global nature of the data constraint.

Another issue that often arises in inverse problems is that of nonuniqueness (see *Inverse Theory, Singular Value Decomposition*). In this case it is not possible to fully constrain the unknowns from the data. The model is unbounded and so best data fit solutions do not exist, and extra assumptions or independent information must be introduced to achieve a single optimal solution. In linearized inversion some form of regularization is used. An example is damping a solution back to some reference set of values (or model) (Aster et al., 2005). It is well known that in this case the details of the solution depend on the nature of regularization used. In addition, uncertainty estimates produced by linearized theory often reflect the choice of regularization. Typically, the least well-constrained components of the solution require the most regularization and resulting uncertainty estimates are severe underestimates of the real errors (Aster et al., 2005 for an example) potential leading to overconfidence in the results.

Bayesian inference

An alternative approach to inversion is Bayesian inference. Many textbooks and review papers are available. Discussions within a geophysical context can be found in Tarantola and Valette (1982), Duijndam (1988a, b), Sambridge and Mosegaard (2002), and Mosegaard and Sambridge (2002). In Bayesian inference, all information on the unknowns is represented in terms of probability density functions (PDF). Within this framework it is accepted that all inference is relative. What one learns from the data gets added to what is known prior to collecting the data and represented in terms of an a posteriori PDF. The most commonly used form of Bayes' rule is given below

$$p(m|d) = kp(d|m)p(m) \tag{1}$$

where $p(m|d)$ is the PDF of the model vector, m, containing the unknowns, given the data vector, d, containing the data; $p(d|m)$ is the likelihood function measuring the probability of the data, d, being observed given the model m; $p(m)$ is the a priori PDF on the model (which is known or assumed about m before the data are

collected), and k is a constant of proportionality. In a Bayesian framework, all information on the unknown variables in the model is represented by the posterior PDF, $p(m|d)$ and one usually sets about trying to generate an ensemble of candidate solutions to the inverse problem whose density is distributed according to this function. This is termed sampling the posterior PDF.

The posterior PDF is the product of the likelihood and the prior PDF. Only the former contains the data vector, d. The likelihood increases as the model fits the data better relative to the noise in the data. The form of the likelihood depends on the statistical character of the data errors. A simple example is a multidimensional Gaussian function characterized by a mean and a covariance matrix, both of which are usually known or assumed. The prior PDF represents information known about the model before collecting the data represented in a probabilistic manner, and may take a variety of forms. Again a multidimensional Gaussian is the most simple, but rarely is real information in this convenient form. Prior PDFs can be the most controversial component of Bayesian inference as there is always a degree of subjectivity in any choice, and the only way to represent no prior information is to not have a prior PDF. Comparisons of Bayesian and alternate approaches can be found in Malinverno and Parker (2005).

We see then that instead of seeking a single optimal solution, in a Bayesian framework many samples are sought. Assessment of the constraints placed on the model is achieved by examining collective properties. Typically, this is done by plotting the distribution of samples as a function of one or more subsets of unknowns, calculating credible intervals to represent uncertainty and covariance matrices to examine the trade-offs between parameters.

The main task to be carried out is then to generate random samples that follow the multidimensional posterior PDF $p(m|d)$arising from the inverse problem. McMC methods are practical tools for dealing with complicated probability distributions. Used correctly they result in (quasi)-independent samples whose density follows any target PDF. They have been the subject of much research in fields from Theoretical Physics to Computational statistics. For summaries, see Smith (1991), Smith and Roberts (1993), and Bernardo and Smith (1994). Below we describe the McMC method briefly and provide a simple illustrative example.

Markov chain Monte Carlo

Fixed dimension approach

McMC can be regarded as a combination of random Monte Carlo sampling and a Markov chain random walk strategy around the model space. The aim is to produce an ensemble of models from a probability distribution, that is, the posterior PDF, using only function evaluations. The basic approach was developed from the work of Metropolis et al. (1953), placed in a Bayesian framework by Hastings (1970), and a useful overview is given in Gilks et al. (1996). The practical applications lagged behind theoretical developments as a consequence of the need for many simulations. However, the increase in computing power over the last 15 years or so has led to a rapid increase in use of this methodology in geophysics and other fields of Earth Sciences (e.g., Mosegaard and Tarantola, 1995; Malinverno, 2002; Sambridge et al., 2006; Gallagher et al., 2009).

The algorithm is as follows: first we choose an initial model from the prior distribution, and calculate its likelihood. Then we generate a new model by making a random perturbation (Monte Carlo) to the current model. This new model is known as the proposed model and depends only on the values of the current model (Markov chain). The final stage is to decide whether we replace the current model with the proposed model, or stay at the current model and repeat the whole process. This important step is determined from the acceptance criterion, which is defined below

$$\alpha = \min\left\{1, \frac{p(m')p(d|m')q(m|m')}{p(m)p(d|m)q(m'|m)}\right\} \tag{2}$$

where m' and m are the proposed and current models, respectively, $q(a|b)$ is the probability of proposing model a, given a current model b, and the other distributions are as defined earlier. The decision to accept or reject a proposed model is made by comparing the value of α (which is always between 0 and 1) to a uniform (between 0 and 1) random number, u. If $u < \alpha$ then we replace the current model with the proposed model, if not we discard the proposed model and stay at the current model. We then continue the sampling process (perturb the new current model and so on) for many iterations.

The choice of the proposal function is not critical to the correctness of the sampler, but does affect the efficiency, performance, and convergence. A typical choice might be a normal distribution, centered on the current model, and then we need to tune the performance through the scale parameter of this distribution (e.g., the variance). If we choose too small a scale parameter, the proposed model will be very similar to the current model, their likelihoods will be similar, and we will almost always accept the proposed model. If we choose too large a scale parameter, the proposed model will tend to be very different to the current model, and lead to large changes in the likelihood, which are more likely to be rejected. In practice, both situations mean that we tend to move slowly around the model space. The proposal functions need then to be tuned for particular problems to achieve a reasonable balance between accepting and rejecting the proposed models. A reasonable rate of acceptance is around 30–40%. Generally, we can choose proposal functions that are symmetrical, so that $q(a|b) = q(b|a)$, so these terms cancel out in the acceptance criterion. Also, if we choose uniform prior distributions, then the prior terms also cancel. The acceptance criterion then reduces to the original Metropolis et al. (1953) algorithm.

After an initial period sampling (known as the burn-in), the current model from each iteration is taken as representing a sample from the posterior distribution (the sampling chain is then stationary). If the model space has N dimensions, and we are interested in the distribution on one of the parameters, m_i, for example, then formally we need to solve the following integral

$$p(m_i) = \int p(m_i, m_j) dm_j; \quad j = 1, .., i-1, i+1, ..., N \tag{3}$$

that is we need to integrate out the variation in all parameters except m_i. This is known as marginalizing and $p(m_i)$is the marginal probability distribution of m_i. Using the McMC samples, we can just plot all value of m_i as a histogram as the sampling effectively deals with the integration.

Also, it is straightforward to calculate estimates of the expected (or average) value for any parameter. Formally, we have the expected value for parameter m_i defined as

$$E(m_i) = \int p(m_i) m_i dm_i \tag{4}$$

Using the McMC samples, we simply average over all the samples accepted for that parameter, that is,

$$E(m_i) = \frac{1}{N_a} \sum_{j=1}^{N_a} m_i^j \tag{5}$$

where N_a is the number of samples accepted (post-burn-in) for model parameter m_i

Transdimensional approach

A major issue concerning most inverse problems, and the approaches used to solve them, is how best to balance the twin desires of fitting the observations and avoiding introduction of unjustified complexity in the resulting models. Green (1995) introduced a transdimensional form of McMC (referred to as Reversible Jump), in which the inversion procedure involves the inference of the model complexity (see also birth–death McMC Geyer and Møller, 1994). For finite dimension models (with a fixed number of unknowns) this then typically becomes a question of determining the dimension of the model. If we are dealing with two models with dimensions k and k', then acceptance criterion can be written as

$$\alpha = \min\left\{1, \frac{p(\mathrm{k}')p(m'|\mathrm{k}')p(d|m', \mathrm{k}')q(m|m')}{p(\mathrm{k})p(m|\mathrm{k})p(d|m, \mathrm{k})q(m'|m)}\right\} \tag{6}$$

Here we separate the prior on the number of dimension, $p(k)$, from the model parameter prior, $p(m|k)$. The proposal function $q()$ becomes more complex as we now want to propose models with different dimensions. Moreover, we need to allow for the transformation from one model to another to ensure that theoretical probability requirements are maintained. In dealing with a situation where we are simply increasing or decreasing the number of parameters, we can write

$$\alpha = \min\left\{1, \frac{p(\mathrm{k}')p(m'|\mathrm{k}')p(d|m', \mathrm{k}')g(u^{k'})}{p(\mathrm{k})p(m|\mathrm{k})p(d|m, \mathrm{k})g(u^{k})} |\mathrm{J}|\right\} \tag{7}$$

Here $u^{k'}$ and u^{k} are vectors of random numbers of length r' and r, respectively, and used to transform from one model to another, such that $r+k = r'+k'$, and $g(.)$ is the probability distribution used to generate these random numbers. The term $|J|$ is the Jacobian, and allows for the transformation between the two models, that is,

$$|\mathrm{J}| = \frac{\partial(m', u^{k'})}{\partial(m, u^{k})} \tag{8}$$

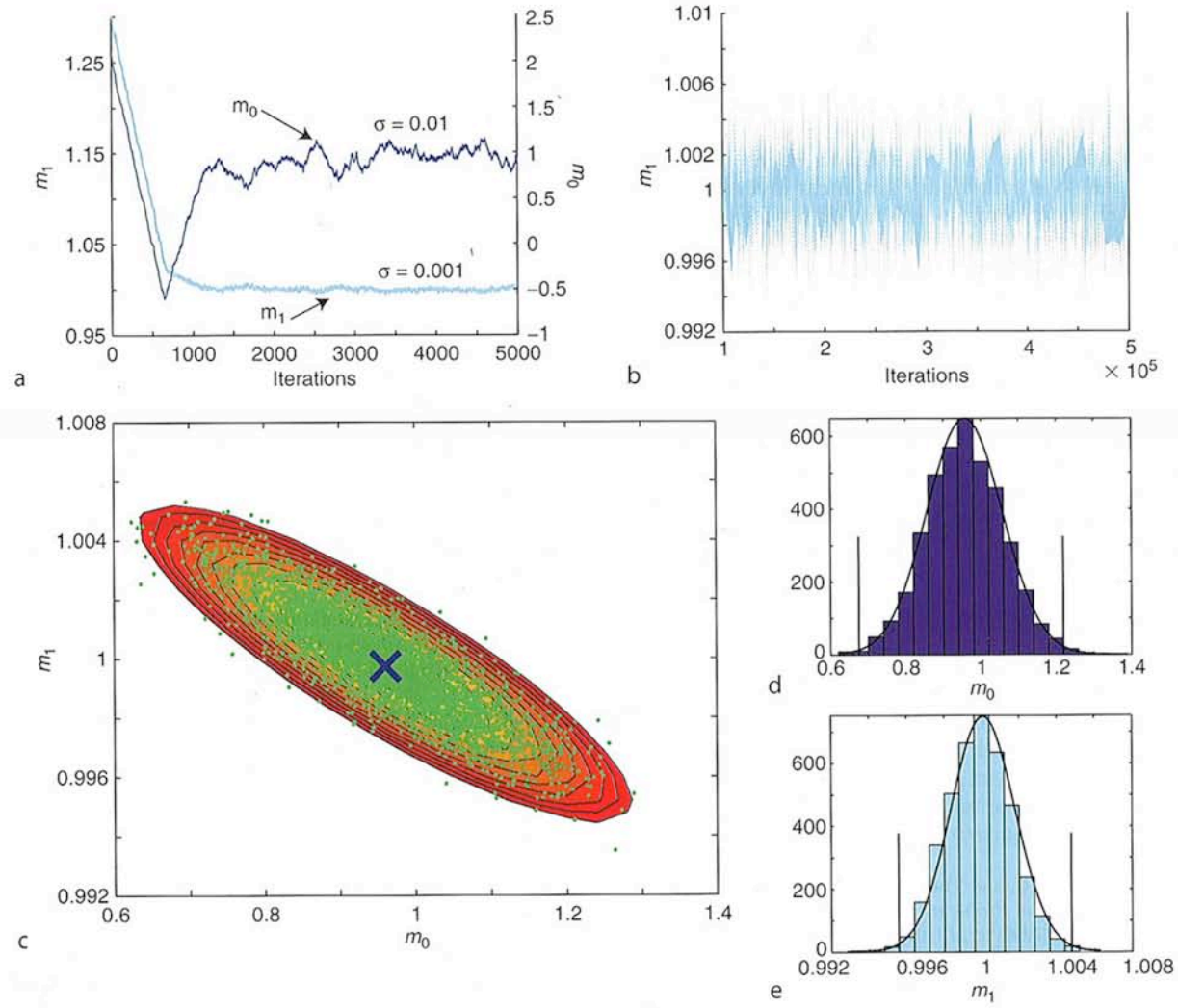
The last equation for α is actually a general form for the acceptance criterion, although for fixed-dimensional problems the Jacobian is generally 1, and the proposal functions are of the form as described earlier. The details of the reversible jump acceptance criterion are discussed in more detail by Green (2003), Malinverno (2002), and Sambridge et al. (2006) and examples of the implementation algorithms for variable dimension problems are given in Jasra et al. (2006), Bodin and Sambridge (2009), Charvin et al. (2009), and Hopcroft et al. (2009). One important characteristic of the Bayesian transdimensional formulation is that it is naturally parsimonious. For two models that fit the data equally well, it will tend to favor simpler models over complex ones as a consequence of the posterior probability effectively being penalized through the addition of more terms to the prior distribution.

A simple example

To demonstrate McMC in action, we choose a simple two parameter linear regression problem, that is,

$$y_i = m_0 + m_1 x_i + \varepsilon_i, i = 1, ..., N \tag{9}$$

where m_0 and m_1 are the model parameters, y is an observed/measured value, and ε is the data error. In a Bayesian formulation, this problem has an analytical solution for a uniform prior on the model parameters, assuming the data error is known (Lee, 1989, p. 180). We chose $m_0 = m_1 = 1$ and generated 100 synthetic data (y) for random values of x between 0 and 100, and added noise ($\varepsilon = 0.5$). We used Gaussian proposal distributions, with different scale parameters ($\sigma_{m_0} = 0.01$, and $\sigma_{m_1} = 0.001$). In Figure 2 we show the sampling for the two parameters, starting from a randomly selected model. Figure 2a shows that the sampler has not reached the stationary state until at least 1,500–2,000 iterations. Figure 2b shows the sampling for parameter m_1 for later iterations which is clearly stationary. The sampling resembles a white noise spectrum about the mean (or expected) value, lacking any internal structure as a function of



Inverse Theory, Monte Carlo Method, Figure 2 (**a**) Intitial 5,000 iterations for sampling of two parameters for the linear regression problem. (**b**) Post-burn-in iterations for parameter m_1. (**c**) The *green points* show the post-burn-in sampling, and the contours are the log likelihood function. The *blue cross* is the best solution (equivalent to the analytical maximum likelihood or least squares solution). (**d**) Marginal distribution for parameter m_0. The histogram is constructed from the post-burn-in accepted samples, and the curve is the analytical solution for the marginal distribution. The two vertical bars mark the upper and lower bounds on the 95% credible interval. (**e**) As (**d**) but for parameter m_1.

iteration. It is also clear that the sampler manages to move toward the upper and lower extreme values of the parameter range (determined by the posterior PDF and the proposal function scale parameter). These are diagnostic (but qualitative) characteristics of stationarity.

Figure 2c shows the 2-D distribution of samples of the post-burn-in accepted samples (here we thinned the chain taking every 100th sample), compared to the log likelihood function (which is proportional to the log of the posterior distribution as we use uniform priors). The density of the sampling increases around the high likelihood values, but there are still some samples from the lower likelihood regions. Figure 2d and e shows the marginal distributions for the two parameters as a frequency histogram, and also the analytical solutions (scaled to the same peak height). These are constructed simply by taking all the accepted values for a given model parameter, as the sampler effectively integrates out the other parameters. Also shown are the 95% credible interval ranges for each parameter. These are constructed by sorting all the samples for a given parameter in ascending order, and finding the indices for the credible values such that 2.5% of the samples are less than the lower credible value, and 2.5% of the values are greater than the upper credible value. Again, we can see that the sampler has managed to distribute itself across the distribution according to the posterior probability and these histograms are good representations of the marginal distributions.

Summary

Monte Carlo sampling, relying on random numbers, has been used in Geophysics for over 40 years, although the increase in computing power has seen a commensurate increase in applications in the last 15 years or so. This approach avoids the use of gradients, is robust to local minima, and so is suitable for nonlinear inverse problems which often have complex misfit (or fitness) surfaces in high dimensions. McMC, particularly when used in a Bayesian formulation, provides a means of sampling a model space according to the (unknown) posterior distribution for the model parameters. Transdimensional (or reversible jump) Markov chain Monte Carlo generalizes this approach to allow models of different dimensions to be considered, and provides a means of choosing between models of differing complexity. Quantifying the posterior distribution with McMC is then a solution to the inverse problem and various types of inference can be made from this distribution (e.g., expected values, marginal distributions, credible intervals) to characterize the model space.

Acknowledgments

We would like to thank FAST (French-Australia Science and Technology exchange program) for their support during the preparation of this article. This project is supported by the Commonwealth of Australia under the *International Science Linkages program.*

Bibliography

Aster, R., Borchers, R., and Thurber, C. H., 2005. *Parameter Estimation and Inverse Problems*. International Geophysics Series, 90. Amsterdam: Elsevier.

Backus, G. E., and Gilbert, J. F., 1967. Numerical applications of a formalism for geophysical inverse problems. *Geophysical Journal of the Royal Astronomical Society*, **13**, 247–276.

Backus, G. E., and Gilbert, J. F., 1968. The resolving power of gross Earth data. *Geophysical Journal of the Royal Astronomical Society*, **16**, 169–205.

Backus, G. E., and Gilbert, J. F., 1970. Uniqueness in the inversion of inaccurate gross Earth data. *Philiosophical Transactions of the Royal Society of London A*, **266**, 123–192.

Bernardo, J. M., and Smith, A. F. M., 1994. *Bayesian Theory*. Chichester: Wiley.

Bodin, T., and Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, **178**, 1411–1436.

Charvin, K., Gallagher, K., Hampson, G., and Labourdette, R., 2009. A Bayesian approach to infer environmental parameters from stratigraphic data 1: methodology. *Basin Research*, **21**, 5–25.

Duijndam, A. J. W., 1988a. Bayesian estimation in seismic inversion part I: principles. *Geophysical Prospecting*, **36**, 878–898.

Duijndam, A. J. W., 1988b. Bayesian estimation in seismic inversion part II: uncertainty analysis. *Geophysical Prospecting*, **36**, 899–918.

Gallagher, K., Charvin, Nielsen, S., Sambridge, M., and Stephenson, J., 2009. Markov chain Monte Carlo (McMC) sampling methods to determine optimal models, model resolution and model choice for Earth science problems. *Marine and Petroleum Geology*, **26**, 525–535.

Geyer, C. J., and Møller, J., 1994. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, **21**, 369–373.

Gilks, W. R., Richardson, S., and Spiegalhalter, D. J., 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Green, P. J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

Green, P. J., 2003. Trans-dimensional McMC. In Green, P. J., Hjort, N., and Richardson, S. (eds.), *Highly Structured Stochastic Systems*. Oxford Statistical Sciences Series, Oxford: Oxford University Press, Chap. 6, pp. 179–196.

Hammersley, J. M., and Handscomb, D. C., 1964. *Monte Carlo Methods*. London: Chapman & Hall.

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Hopcroft, P., Gallagher, K., and Pain, C. C., 2009. A Bayesian partition modelling approach to resolve spatial variability in climate records from borehole temperature inversion. *Geophysical Journal International*, **178**, 651–666.

Jasra, A., Stephens, D. A., Gallagher, K., and Holmes, C. C., 2006. Analysis of geochronological data with measurement error using Bayesian mixtures. *Mathematical Geology*, **38**, 269–300.

Kennett, B. L. N., Brown, D. J., Sambridge, M., and Tarlowski, C., 2003. Signal parameter estimation for sparse arrays. *Bulletin. Seismological Society of America*, **93**, 1765–1772.

Lee, P. M., 1989. *Bayesian Statistics: an Introduction*. New York/ Toronto: Edward Arnold.

Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a non linear geophysical problem. *Geophysical Journal International*, **151**, 675–688.

Malinverno, A., and Parker, R. L., 2005. Two ways to quantify uncertainty in geophysical inverse problems. *Geophysics*, **71**, 15–27.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., 1953. Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, **21**, 1087–1092.

Mosegaard, K., and Sambridge, M., 2002. Monte Carlo analysis of inverse problems. *Inverse Problems*, **18**, R29–R54.

Mosegaard, K., and Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, **100**, 12431–12447.

Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithm -I. Searching a parameter space. *Geophysical Journal International*, **138**, 479–494.

Sambridge, M., and Mosegaard, K., 2002. Monte Carlo methods in Geophysical inverse problems. *Reviews of Geophysics*, **40**, 3.1–3.29.

Sambridge, M., Gallagher, K., Jackson, A., and Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence. *Geophysical Journal International*, **167**, 528–542.

Smith, A. F. M., 1991. Bayesian computational methods. *Philiosophical Transactions of the Royal Society of London A*, **337**, 369–386.

Smith, A. F. M., and Roberts, G. O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, **55**, 3–23.

Tarantola, A., and Valette, B., 1982. Inverse problems = quest for information. *Journal of Geophysics*, **50**, 159–170.

Cross-references

Inverse Theory, Artificial Neural Networks
Inverse Theory, Global Optimization
Inverse Theory, Linear
Inverse Theory, Singular Value Decomposition