Sampling informative/complex a priori probability distributions using Gibbs sampling assisted by sequential simulation

Thomas Mejer Hansen, Klaus Mosegaard, and Knud Skou Cordua[1]

[1]Center for Energy Resources Engineering, Department of Informatics and Mathematical Modelling, DTU, Richard Petersens Plads, Building 321, 2800 Lyngby, Demmark. E-mail: tmeha@imm.dtu.dk

Abstract

Markov chain Monte Carlo methods such as the Gibbs sampler and the Metropolis algorithm can be used to sample the solutions to non-linear inverse problems. In principle these methods allow incorporation of arbitrarily complex a priori information, but in practice current methods allow only relatively simple priors to be used. We demonstrate how sequential simulation can be seen as an application of the Gibbs sampler, and how such a Gibbs sampler assisted by sequential simulation can be used to perform a random walk that generates realizations of a relatively complex random function. We propose to combine this algorithm with the Metropolis algorithm in order to obtain an efficient method for sampling posterior probability densities for non-linear inverse problems.

Keywords: Monte Carlo, prior, sequential simulation, inverse problem.

1. INTRODUCTION

Consider a typical forward problem

$$d = g(m) \quad (1)$$

where a function *g* relates a subsurface model *m* to observational data *d*. Inverse problem theory deals with the problem of inferring properties of *m* from a specific dataset *d*, using eqn. (1) and some prior information on *m*. Tarantola (2005) and Mosegaard (2006) formulated a probabilistic approach to solving inverse problems where a priori information is described by an a priori probability density function (pdf), $\rho_M(m)$, and the data fit associated to a given model is given by a likelihood function, $L(m)$. The solution to such an inverse problem is the a posteriori probability density, which is proportional to the product of the prior and the likelihood:

(2) $$\sigma_M(m) = k\,\rho_M(m)\,L(m)\ ,$$

where *k* is a normalization factor. Hansen et al. (2006) propose an efficient, non-iterative approach using sequential simulation to generate samples of the a posteriori pdf in case *g* is a linear function and both $\rho_M(m)$ and $L(m)$ can be described by Gaussian statistics. Hansen and Mosegaard (2008) relax the Gaussian assumption of this approach, which allows for non-Gaussian a priori distributions.
In practice, *g* is often a nonlinear operator, and $L(m)$ and $\rho_M(m)$ are non-Gaussian, which results in an non-Gaussian a posteriori pdf. Mosegaard and Sambridge (2002) summarize and discuss a number of Monte Carlo based methods for sampling the solution to such problems. Among these we find the rejection sampler, the Gibbs sampler, and the Metropolis algorithm. Each of these methods is guaranteed to sample the a posteriori pdf asymptotically, although the computational efficiency may differ significantly.
They allow an arbitrarily complex noise model and arbitrarily complex a priori information to be used, but they differ in the way the content of the a priori model is presented to the algorithm. A short description of each of these methods, and their demands on the a priori model, is given here.

Rejection sampler

The rejection sampler is perhaps the simplest method for sampling the posterior probability density function, $\sigma_M(m)$ (eqn. (2)). It allows inclusion of complex a priori information, and any black box that generates independent realizations from the a priori probability density function can be used. Rejection sampling works by filtering a list of independent realizations of the a priori model. Each proposed model is accepted with probability

(3) $$p_{acc}(m_{propose}) = L(m_{propose})/M$$

where *M* is larger than (or equal to) the maximum likelihood of all the proposed models. In many cases the maximum likelihood is not known, and one must set *M* to a large value. For large-dimensional problems this typically causes the acceptance probability p_{acc} to be very small, and hence the algorithm to be very inefficient.

Metropolis algorithm

The Metropolis algorithm is a Monte Carlo sampling method based on Markov chains (Metropolis & Ulam, 1949). Mosegaard and Tarantola (1995) describe a generalized Metropolis algorithm that allows analysis of non-linear inverse problems with complex a priori information. The prior information must be quantified in such a way that samples of the a priori probability density, which are (often) a perturbation of the previous sample, can be obtained. Furthermore, one must be able to control the exploratory nature of the prior sampler (i.e. the step length between two successive prior samples) in order to control the efficiency of the algorithm.
Each iteration of the Metropolis algorithm, starting in model m_n consists of two stages: a) exploration and b) exploitation. In the exploration stage, one step of a random walk, sampling the prior, is performed. In other words, an unconditional realization m_{n+1} of the a priori pdf $\rho_M(m)$ in the vicinity of m_n is generated. This is followed by the exploitation stage where the

likelihood of the proposed model is evaluated. If $L(m_n)$ and $L(m_{n+1})$ is the likelihood of m_n and m_{n+1}, respectively, m_{n+1} is accepted with probability:

$$(4) \qquad P_{accept} = \begin{cases} 1 & if\ L(m_{n+1}) > L(m_n) \\ L(m_{n+1})/L(m_n) & otherwise \end{cases}$$

If m_{n+1} is accepted, m_{n+1} becomes the current model. Otherwise m_n remains the current model. Performed iteratively this algorithm will sample the a posteriori pdf, in the sense that output models occur with a frequency proportional to their a posteriori probability.
The computational efficiency of the Metropolis algorithm depends on the structure of the posterior distribution to be sampled. In addition, it is strongly dependent on the exploratory nature (i.e. step length) of the prior sampler. No theoretically correct step length can be found, but Gelman et al. (1996) suggest that a step length giving rise to an acceptance ratio of the Metropolis sampler of about 25-50% is reasonable. In any case, one must be able to adjust the step length in order to successfully apply a specific method to sample the prior.

Gibbs sampler

In each step of the Gibbs sampler (Geman and Geman, 1984), a model parameter m_i is selected at random. Then the conditional probability distribution of m_i, given that the rest of the model parameters are held constant, is computed. Finally, a realization of m_i is drawn from the conditional distribution.
An important property of the Gibbs sampler is that no models are rejected, as is the case for the Metropolis sampler. In most implementations, the main computational task of applying the Gibbs sampler is to compute and draw realization from the conditional pdf. However, it is important to note that computing the local conditional distribution is not a requirement. Any method that is able to generate realizations from the conditional distribution will suffice.
A number of conditions must be satisfied in order to ensure that the Metropolis algorithm and the Gibbs sampler sample the desired distribution. First, *aperiodicity* and *irreducibility* must be satisfied to ensure that the algorithm has a unique *equilibrium distribution* (Mosegaard and Sambridge, 2002). In addition, the desired distribution must be the equilibrium distribution of the algorithm. This is satisfied in a simple way if each pair of neighboring sample points are in *detailed balance*: The probability that a jump takes place from model m_k, to m_l must be equal to the probability that a jump takes place from model m_l to m_k:

$$(5) \qquad P(m_k \rightarrow m_l \mid m_k)\ \rho_M(m_k) = P(m_l \rightarrow m_k \mid m_l)\ \rho_M(m_l)$$

In this manuscript we will focus on ways to quantify complex a priori information such that it can be used with the rejection sampler, the Metropolis algorithm, and the Gibbs sampler. We will demonstrate how geostatistical algorithms, based on sequential simulation, are capable of simulating geological reasonable structures and therefore are suitable for quantifying complex a prior information. These algorithms can be used directly with the rejection sampler. We will show that the Gibbs sampler and sequential simulation are closely related. Specifically we will demonstrate an application of the Gibbs sampler that will enable generating realizations from any stochastic model that can be simulated using sequential simulation. Further we will demonstrate that this sampling algorithm honors detailed balance, such that it will actually sample the stochastic model intended. We suggest using this method, which we refer to as

sequential Gibbs sampling, to control the exploratory nature (i.e. step length) of the Metropolis algorithm. This strategy not only leads to the formulation and solution of inverse problems with complex a priori information, but can also have dramatic effect on the computational efficiency of the Metropolis sampling algorithm.

2 QUANTIFYING PRIOR INFORMATION USING GEOSTATISTICS

Generally speaking geostatistics is an application of random functions to describe spatial phenomena, typically in form of spatial variability in earth models. Geostatistical simulation algorithms have been developed to efficiently generate realization of a number of random function models. Geostatistical simulation algorithms can be divided into two groups, where the underlying random function model is based on 2-point and multiple-point statistics, (Guardiano and Srivastava,1993, Strebelle, 2002). 2-point based geostatistical algorithms take into account spatial variability between sets of two data locations. In case the distribution of the model parameters is Gaussian, one can completely define the underlying random function model using a Gaussian pdf:

$$\rho_M(m) = c\ \exp(-0.5(m-m_{prior})^t\, C^{-1}_{M_{prior}}\,(m-m_{prior})), \quad (6)$$

where m_{prior} is the a priori mean and $C^{-1}_{M_{prior}}$ is the a priori covariance model.

Multiple-point-based geostatistical models have no parametric description. Instead multiple point statistics is inferred from training images. The methodology was initially proposed by Guardiano and Srivastava (1993). Strebelle (2002) developed the first computationally feasible algorithm for categorical training images. Zhang et al. (2006) and Wu et al. (2008) suggested another multiple point based algorithm where patterns from a continuous or categorical training image are used to generate stochastic realizations with features from the training image. Using these techniques one can generate realization of random function models that reproduce geologically realistic spatial variability. State of the art implementations of these algorithms are available through SGeMS (Remy et al., 2008).

2.1 Sequential simulation

Consider N points $\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_N$ and a random field $Z(\mathbf{u})$ describing the interdependence between values of a physical or geological property measured at points $\mathbf{u}_i$. Then, one realization $\boldsymbol{z}(\mathbf{u})$ of the vector

$$\boldsymbol{Z}(\mathbf{u}) = (Z(\mathbf{u}_1), Z(\mathbf{u}_2), \ldots, Z(\mathbf{u}_n), Z(\mathbf{u}_{n+1}), \ldots Z(\mathbf{u}_N)) \quad (7)$$

can be simulated using sequential simulation as follows: In step *i,* visit location $\mathbf{u}_i$ and draw a random value $z_i = \boldsymbol{z}(\mathbf{u}_i)$ from $\boldsymbol{Z}(\mathbf{u}_i)$, using the conditional probability density function

$$f_Z(z_i \,|\, z_1 \ldots z_{i-1}), \quad (8)$$

where $f_Z(z_1 \dots z_N)$ is the joint probability density of the components of **Z**(**u**). That the above procedure will actually sample $f_Z(z_1 \dots z_N)$ follows from the identity $f(s|t)f(t) = f(s,t)$, which (in the general multivariate case) yields

$$f_Z(z_{n+1} \dots z_N | z_1 \dots z_n) = \begin{matrix} f_Z(z_{n+1} | z_1 \dots z_n) \\ f_Z(z_{n+2} | z_1 \dots z_{n+1}) \\ \vdots \\ f_Z(z_N | z_1 \dots z_{N-1}) \end{matrix} \quad (9)$$

When all locations have been visited once, a realization of **Z**(**u**) is generated. Thus to apply sequential simulation one must a) build a local conditional pdf (conditional to the previously simulated data), and b) draw a realization of this local pdf. Considerable effort has been made in the geostatistical community to efficiently compute conditional probability density functions based on the 2-point and multiple-point stochastic models as described above.

2.2 Sequential Gibbs sampling

Consider a known realization **z**(**u**) of the random function **Z**(**u**) obtained using sequential simulation. If we now, at random, select a model parameter, $z_i = \boldsymbol{z}(\mathbf{u}_i)$, compute the local conditional pdf

$$f_Z(z_i | z_1, z_2, \dots, z_{i-1}, z_{i+1}, \dots, z_n) \quad (10)$$

and draw a value from it, we get a new realization of the random field **Z**(**u**). If this is repeated iteratively, this will be an application of the Gibbs sampler (Geman and Geman, 1984). The cost of using the Gibbs sampler is that one must be able to generate a realization of the local conditional pdf, which can be done very effectively using methods developed for sequential simulation. We refer to this combination of sequential simulation and Gibbs sampling as sequential Gibbs sampling.

The sequential Gibbs sampler can be used as a prior sampler for the generalized Metropolis algorithm (Mosegaard and Tarantola, 2005). However, in order to control the computational efficiency of this algorithm, some flexibility of the amount of perturbation (i.e. some control of the 'step-length' of the prior sampler) is needed. We suggest considering not just one model parameter at each step of the Gibbs sampler, but a subset *U* of model parameters. Assuming that the model parameters we wish to update belong to such a subset, we need to generate a realization of the conditional pdf

$$f_Z(z_{i \in U} \mid z_{i \notin U}) \quad (11)$$

Recall that we do not need to explicitly calculate the complete conditional distribution in eqn. (11), but only to generate a realization from it. In order to do this we can make use of the sequential simulation approach of eqn. (9), which involves computing only the conditional probability density function for each model parameter in *U* in random order.

In this way we have designed a Gibbs sampler that efficiently samples a prior pdf by utilizing the technique of sequential simulation to draw values from the local conditional pdf. At the

same time this technique provides a means of controlling the step length of the prior sampling by varying the number of model parameters in the subset

Detailed Balance
Mosegaard and Sambridge (2002) demonstrate that the Gibbs sampler satisfies detailed balance. As the Sequential Gibbs sampler is an application of the Gibbs sampler, this property is also ensured for the sequential Gibbs sampler. This means that the random walk performed by running the sequential Gibbs sampler will asymptotically sample the same random function as a traditional sequential simulation.

Algorithm for sequential Gibbs sampling
Implementing the sequential Gibbs sampler amounts to implementing a Gibbs sampler which, in each iteration, calculates a realization of the conditional probability density function associated to a specific subset of model parameters using sequential simulation:

1. Select a region (i.e. subset) in the physical space and regard all model parameters associated to this area as unknowns. The rest of the model parameters are considered known (and fixed).
2. Perform sequential simulation of the unknown parameters conditioned to the known parameters. This generates a new model, which is also a realization of the prior model. This step is identical to drawing a value from the conditional probability density function in eqn. (11).
3. Use the new model as the starting model and go to 1.

Such an algorithm was proposed by Hansen et al. (2008). They did, however, not make the link to the Gibbs sampler, and they provided no proof that the resulting algorithm samples an equilibrium distribution, nor that such an equilibrium distribution would in fact be the requested a priori model.
As already mentioned, the size of the region of model parameters that is regarded as unknown (in step 1 in the sequence above) can be used to control the step length of the sequential Gibbs sampler. By regarding only a single model parameter as unknown results in a model m_{i+1} which is highly related to the original model m_i . On the contrary, by regarding all model parameters as unknowns leads to m_{i+1} which is statistically independent of m_i .

2.3 Gradual deformation

Techniques that allow a gradual deformation between two realizations of random function have been developed for both 2-point based Gaussian random function models (the Gradual Deformation Method(GDM) by Hu, 2000; Le Ravalec et al., 2000) and multiple-point based random function models (the Probability Perturbation Method (PPM) by Caers and Hoffman, 2006). The use of Gradual Deformation has mainly been as part of an optimization algorithm for data calibration, where iterative gradual deformation have been used to gradually change a starting model until the forward response from the model matches observed data to some satisfactory degree (Caers and Hoffman, 2006). It has been suggested that running such an application several times, generating a set of models that all fit the data, can be used to describe posterior uncertainty. This is however not the case. The resulting model variability of such an approach reflects the choice of optimization algorithm and the level chosen for

acceptable data fit. An effort to address this issue was made by Le Ravalec-Dupin and Noetinger (2002).
If GDM and PPM honors microscopic reversibility they could, however, be used to sample the prior.

3 CONCLUSIONS

We have demonstrated that the Gibbs sampler and the method of sequential simulation are closely related. We have proposed an efficient sampling algorithm - sequential Gibbs sampling combining these two methods. Sequential Gibbs sampling can sample random function models based on 2-point as well as multiple-point based statistics. We have shown that this algorithm satisfies microscopic reversibility and that the equilibrium distribution is in fact the prior probability density. The sequential Gibbs sampler can be run with arbitrary 'step lengths'. The longest step length (where all parameters are updated in one step) result in a new realization from the a prior distribution which in statistically independent to the previous model. A step length of zero returns the same realization as the previous model. This makes the sequential Gibbs sampler well suited as a method for sampling a complex prior in connection with the Metropolis sampler.
Utilization of the sequential Gibbs sampler will allow relatively efficient analysis of the solution to nonlinear inverse problems with complex a priori information.

ACKNOWLEDGEMENT

The present work was sponsored by DONG Energy as part of the research project 'Risk assessment and multiple scenario generation from seismic and geological data'.

REFERENCES

Caers, J. & Hoffman, T., 2006. The probability perturbation method: A new look at bayesian inverse modeling, *Mathematical Geology*, **38**(1), 81 – 100.
Gelman, A., Roberts, G., & Gilks, W., 1996. Efficient metropolis jumping hules, in *Bayesian Statistics 5,*, pp. 599–608, Clarendon press, Oxford.
Geman, S. & Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intell*, **6**, 721–741.
Guardiano, F. & Srivastava, R., 1993. Multivariate geostatistics: beyond bivariate moments, *Geostatistics-Troia*, **1**, 133–144.
Hansen, T. & Mosegaard, K., 2008. VISIM: Sequential simulation for linear inverse problems, *Computers and Geosciences*, **34**(1), 53–76.
Hansen, T. M., Journel, A. G., Tarantola, A., & Mosegaard, K., 2006. Linear inverse Gaussian theory and geostatistics, *Geophysics*, **71**(6), R101–R111.
Hansen, T. M., Mosegaard, K., & Cordua, K. C., 2008. Using geostatistics to describe complex a priori information for inverse problems, in *VIII International Geostatistics Congress*, vol. 1, pp. 329–338, Mining Engineering Department, University of Chile.
Hu, L. Y., 2000. Gradual deformation and iterative calibration of gaussian-related stochastic models, *Mathematical Geology*, **32**(1), 87–108.
Metropolis, N. & Ulam, S., 1949. The Monte Carlo Method, *Journal of the American Statistical Association*, **44**(247), 335–341.

Mosegaard, K., 2006. *Monte Carlo Analysis of Inverse Problems*, University of Copenhagen.
Mosegaard, K. & Sambridge, M., 2002. Monte carlo analysis of inverse problems, *Inverse Problems*, **18**(3), 29–54.
Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *Journal of Geophysical Research*, **100**(B7), 12431–12447.
Ravalec, M. L., Noetinger, B., & Hu, L. Y., 2000. The fft moving average (fft-ma) generator: An efficient numerical method for generating and conditioning gaussian simulations, *Mathematical Geology*, **32**(6), 701–723.
Remy, N., Boucher, A., & Wu, J., 2008. *Applied Geostatistics with SGeMS: A User's Guide*, Cambridge Univ Pr.
Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics, *Math. Geol*, **34**(1), 1–20.
Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.
Wu, J., Boucher, A., & Zhang, T., 2008. A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM, *Computers and Geosciences*, **34**(12), 1863–1876.
Zhang, T., Switzer, P., & Journel, A., 2006. Filter-based classification of training image patterns for spatial simulation, *Mathematical Geology*, **38**(1), 63–80.