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

Hansen, Thomas Mejer; Mosegaard, Klaus; Cordua, Knud Skou

Published in:
IAMG Extended Abstracts

Publication date:
2010

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 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 generating realizations of a relatively complex random function. We propose to combine this algorithm with the Metropolis algorithm 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 equation (1) and some prior information on m. Tarantola (2005) and Mosegaard(2006) formulate a probabilistic approach to solving inverse problems where a priori information is described by the a priori probability density function (pdf) $\rho_M(m)$, and the data fit associated to a given model is given by the likelihood, $L(m)$. 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) $$\sigma_M(m) = \rho_M(m)\, L(m)$$

where k is a normalization factor. In case g is a linear function, and both $\rho_M(m)$ and $L(m)$ can be described by Gaussian statistics, Hansen et al. (2006) propose an efficient, non-iterative approach, using sequential simulation, to generate samples of the a posteriori pdf. Hansen and Mosegaard (2008) relax the Gaussian assumption and allow an a priori model with a non-Gaussian distribution.

In practice, g is often a nonlinear operator, and $L(m)$ and $\rho_M(m)$ are non-Gaussian. 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, $\rho_M(m)$ in 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 the prior probability density can be sampled in the model space. Furthermore, in order to control the efficiency of the algorithm one must be able to control the exploratory nature of the random walk, i.e., one must be able to control its the step length.

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. Assume that the likelihood of m_n and m_{n+1} with respect to the observed data is $L(m_n)$ and $L(m_{n+1})$, respectively. Then m_{n+1} is accepted with probability Paccept:

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

If m_{n+1} is accepted, m_{n+1} becomes the current model. Otherwise the model m_{n+1} is rejected, and m_n remains the current model. Performed iteratively this algorithm will sample the a posteriori pdf, in the sense that models are accepted with a frequency proportional to their a posteriori probability.

The computational efficiency of the Metropolis algorithm of course depends on the complexity of the posterior distribution to sample. In addition, it is strongly dependent on the exploratory nature of the prior sampler, i.e. the 'step length' of the prior sampler. No theoretical 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, for a successful application of a specific choice of method for performing a random walk in the prior model space, one must be able to adjust the exploration rate.

Gibbs sampler

The Gibbs sampler (Geman and Geman, 1984) is - as the Metropolis algorithm - a random walk algorithm that can be used to sample the a posteriori pdf. It differs from the Metropolis algorithm in that at each step in the random walk, a model parameter mi is selected at random. Then the conditional probability density function, given that the rest of the model parameters are held constant, is computed. Finally, a realization of mi is drawn from the conditional pdf.

An important property of the Gibbs sampler is that no models are rejected, as is the case for the Metropolis sampler. The main computational task of applying the Gibbs is most often to actually compute and draw realization from the conditional pdf. Note, that one does not strictly need to compute the local conditional distribution. A method for generating realizations from the conditional distribution will suffice.

In order to ensure that both the Metropolis algorithm and the Gibbs sampler will end up sampling the correct a posteriori distribution, a number of conditions must be satisfied. First, the sampling algorithm must reach an equilibrium distribution. 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:

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

In addition, *aperiodicity* and *irreducibility* must be satisfied to ensure that the equilibrium distribution is actually the a priori distribution $\rho_M(m)$ (Mosegaard and Sambridge, 2002).

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 and capable of simulating geological reasonable structures, can be used to quantify 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 microscopic reversibility, such that it will actually sample the stochastic model intended. We suggest to use this method, which we refer to as *sequential Gibbs sampling*, to handle the exploratory part of the Metropolis algorithm. This 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, guardiano1993multivariate,Strebelle. 2-point based geostatistical algorithms take into account spatial variability between sets of 2 data locations. In case the distribution of the model parameters are Gaussian, one can completely define the underlying random function model using a Gaussian pdf.

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

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 the multiple point statistics are inferred from a training image. 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 implementation of these algorithms are available through SGeMS (Remy et al., 2008).

2.1 Sequential simulation

Consider N sampling points $\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_N$ and a random field $Z(\mathbf{u})$ describing the spatial relations between the sampling points, then one realization of the vector

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

can be simulated as $\mathbf{z}(\mathbf{u}) = (\boldsymbol{z}(\mathbf{u}_1), z(\mathbf{u}_2), \ldots z(\mathbf{u}_N))$ using sequential simulation as follows. Sequentially visit all locations and draw a value $\boldsymbol{z}(\mathbf{u}_i)$ from $\boldsymbol{Z}(\mathbf{u}_i)$, by randomly drawing a value from the conditional probability density function

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

for *i = n+1 ,…, N*. That this is true follows from the identity $f(s\,|t)\,f(t) = f(s,t)$, which (in the general multivariate case) yields

$$\begin{aligned} f_Z(z_{n+1} \ldots z_n \,|z_1 \ldots z_n) &= f_Z(z_{n+1} \,|z_1 \ldots z_n) \\ &= f_Z(z_{n+2} \,|z_1 \ldots z_{n+1}) \\ &\vdots \\ &= f_Z(z_N \,|z_1 \ldots z_{N-1}) \end{aligned} \tag{9}$$

When all locations have been visited one realization of **Z**(**u**) is generated as *z*(**u**). 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 have been made in the geostatistical community to efficiently compute conditional probability density functions, based on the 2-point and multiple stochastic models presented earlier.

2.2 Sequential Gibbs sampling

Recall that the Gibbs sampler, just as the sequential simulation method, is dependent on an efficient method to generate a local conditional probability distribution. Consider **z**(**u**) as a known realization of the random function **Z**(), obtained using sequential simulation. If we now, at random, select a model parameter, z_i , compute the local conditional pdf

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

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 sequential simulation. We refer to this combination of sequential simulation and Gibbs sampling as sequential Gibbs sampling.
If such sequential Gibbs sampling should be used to sample prior information, in conjunction with the generalized Metropolis algorithm, to sample the posterior of inverse problems, then some flexibility of the amount of perturbation is needed, i.e. some control of the 'step-length' of the prior sampler is needed, in order to control the computational efficiency of the generalized Metropolis sampler. We suggest to consider not just one model parameter at each step of the Gibbs sampler, but a set of model parameters. Say the model parameters we wish to update belongs to *I*, then we need to generate a realization of the conditional pdf

(11) $f_Z(z_{i \notin I} \,|z_{i \in I})$

Recall that we do no need to explicitly calculate the complete local conditional distribution in eqn. 11, but only be able generate a realization from it. For 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 *I* in random order.

We suggest to use a Gibbs sampler to sample the prior information quantified by random function models that can be simulated using sequential simulation. For such random function models, sequential simulation can be used to efficiently draw a value from the local conditional pdf, which is needed by the Gibbs sampler.

Microscopic reversibility
Mosegaard and Sambridge (2002) demonstrate that the Gibbs sample respects microscopic reversibility. As the Sequential Gibbs sampler is an application of the Gibbs sampler this, microscopic reversibility is also ensured for the sequential Gibbs sampler. This means that the random walk performed by running the sequential Gibbs sampler will end up sampling exactly the same random function as samples running sequential simulation!

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

1. In the current model m_i, select a region in the model space, and denote all model parameters in this area as unknown, $m_i(u)$. The rest of the model parameters are considered known $m_i(k)$.
2. Perform sequential simulation of $m_i(u)$, conditioned to $m_i(k)$. This generates a new model m_{i+1}, 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. Set $m_i = m_{i+1}$ 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. They provided no proof that the resulting algorithm would sample an equilibrium distribution, nor that such a equilibrium distribution would in fact be the requested a priori model.

2.3 Gradual deformation

Gradual deformation techniques that allow a gradual deformation between two realizations of random function has been developed for both 2-point based Gaussian random function models (Gradual Deformation Method, GDM) (Hu, 2000; Le Ravalec et al., 2000) and multiple point based random function models (Probability Perturbation Method, PPM) (Caers and Hoffman, 2006). The main use of gradual deformation have 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 match 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 variability apparent in such a set of models, reflects the choice of optimization algorithm and the level chosen for acceptable data fit. An effort to address this is issue was made by Le Ravalec-Dupin and Noetinger (2002).
If GDM and PPM honors microscopic reversibility they could though be used to perform a random walk in the a priori probability space. They could also be used in conjunction with the sequential Gibbs sampler to allow a gradual perturbation to the currently visited model, and thus allow for a more detailed control of the perturbation.

3 CONCLUSIONS

We have demonstrated the how the Gibbs sampler and the method of sequential simulation are close related. We have proposed an efficient sampling algorithm combing these two methods called sequential Gibbs sampling, that can perform a random walk in an a priori model parameter space described by random function models based on both 2-point and multiple point based statistics. We have shown that this algorithm honors microscopic reversibility and that the equilibrium distribution is in fact the probability density function described by the a priori choice of random model. The sequential Gibbs sampler can be run with arbitrarily step lengths. The longest step length result in a new realization from the a prior distribution that is uncorrelated to the previous model. A step length of zero returns the same realization as the previous model. This make the sequential Gibbs sampler well suited acting as a method to perform a random walk in the prior model parameter space, which is needed to quantify such complex a priori information for use 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.

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.