Tu P2 02

An Adaptive Sampling Strategy to Accelerate Markov Chains Monte Carlo

H. Hamdi* (University of Calgary), I. Couckuyt (Ghent University-iMinds), M. Costa Sousa (University of Calgary), T. Dhaene (Ghent University-iMinds), C.R. Clarkson (University of Calgary)

Summary

Markov chain Monte Carlo (McMC) is a technique to sample the posterior distributions in order to quantify the uncertainty and update our knowledge about the model parameters. However, this technique is computationally costly. Therefore, it is frequently combined with a cheap-to-evaluate proxy model. Building an efficient and accurate proxy model using one-shot sampling method (such as Latin-Hypercube) is a challenging task as we do not know a prior the best sample locations and the number of required samples to tune our proxy model. In this paper, we present a novel adaptive sampling techniques using LOLA-Voronoi and Expected Improvement techniques to sequentially update our proxy (here ordinary Kriging model) using best sampling locations and density. This algorithm is used to balance between the exploration and exploitation to create an accurate proxy model for the IC fault model's misfit function. The results show a large improvement comparing to the one-shot Latin-Hypercube design. The built proxy is then employed to reduce the computational cost of McMC process to find the posterior distributions of the model parameters.

Introduction

The Bayesian framework is a statistically consistent way to update our knowledge about the model parameters using the observations (Christie et al., 2006). The Bayesian framework is formulated as:

$$f(\mathbf{m}|\mathbf{Y_o}) = \alpha f(\mathbf{m})L(\mathbf{m}|\mathbf{Y_o}) \qquad\qquad 1$$

in which, $\mathbf{m}$ is a set of unknown model parameters, $f(\mathbf{m})$ is our prior belief about the model parameters, $f(\mathbf{m}|\mathbf{Y_o})$ is the posterior distribution, $L(\mathbf{m}|\mathbf{Y_o})$ is the likelihood (representing a mismatch between the numerical model response and the observed data), and $\propto$ is a normalizing constant.

Markov chain Monte Carlo (McMC) (Metropolis et al., 1953) is a statistical technique which can be used to evaluate Equation 1 by generating a chain of dependent samples from a stable (high dimensional) posterior distribution (Privault, 2013; Vrugt, 2016). Each new location (sample or set of parameters) in the chain is obtained by moving from the current location with a probability that only depends on the current location. The dependency between two successive samples in the chain is a controlling factor in McMC (Mariethoz and Caers, 2014).

One of the practical computational issues in performing the McMC is that it requires a large number of costly samples (simulations) to be evaluated. To solve this issue, the surrogate modeling is employed to reduce the computational cost. A surrogate model (e.g. Kriging) is usually built by selecting a mathematical function that is tuned by drawing a few (intelligent) samples from the original system and observing the outcomes.

Traditional experimental design methods such as Latin-Hypercube (Santner et al., 2014) create a certain number of samples in one-shot, then after evaluating the output of all samples a one-time static surrogate model is built. If the system is not very complicated (e.g. linear), only a few samples can be used to build a proxy model to mimic the true response. However, sometimes the true complexity of the underlying function is large and it requires a large number of sample evaluations before an accurate proxy is built. This indicates the issues of under- and over-sampling which arise from the use of one-shot experimental designs. The deficiency in that we do not know *a prior* how many samples to choose can immediately show that such static sampling strategies could provide inefficient and/or inaccurate surrogate models which are of little use in the McMC process. In addition, the one-shot sampling strategies rely only on the input parameters and no information from the output is used to design better sample configurations. Therefore, they are only exploratory designs which fail to exploit the regions of interest with various nonlinearities. Example of traditional sampling methods for building a proxy is that of Yang et al. (2015) who presented a surrogate-based workflow for probabilistic forecasting using Radial Basis Functions.

Adaptive or sequential sampling strategies, on the other hand, analyze the output of each sample and select the next samples in the nonlinear regions of the true objective function or places that multiple optima are located. Therefore, only a limited number of samples are used to define a final surrogate model. Adaptive sampling strategies construct and tune a new surrogate model at each iteration. These approaches use the one-shot methods as starting points and then make a trade-off between exploration and exploitation. One of the more recent powerful Monte Carlo-based methods for adaptive sampling is Local Linear Approximation (LOLA)-Voronoi (Crombecq et al., 2009) or its high dimensional counterpart FLOLA-Voronoi (van der Herten et al., 2015). These methods can balance between the exploration and exploitation and therefore can be suited for optimization or uncertainty analysis using the surrogate modeling approach. The LOLA-Voronoi can even be combined with other optimization-based infill criteria such as Expected-Improvement (Jones, 2001) which finds the next best samples by computing the improvement we expect, compared to the current best value of the objective function. With help of an adaptive sampling method, we can intuitively construct an appropriate *global* proxy model that can replace the misfit function surface. This can facilitate the samplings and accelerate the McMC process.

In this paper, we implement an adaptive sampling method to build a tuned ordinary Kriging model that can be used to generate the posterior distributions of the model parameters for the IC fault model (Tavassoli et al., 2005).

Example: The IC fault model

The IC fault model is a 2D multi-layer faulted cross-sectional model with an injection and a producing well. The wells are separated by a single fault with a variable throw, which controls the flow between the left and the right blocks of the model. The reservoir thickness is 60 ft with six alternating layers of high and low quality facies. The full discussion of the objective function can be found in (Tavassoli et al., 2005). To perform the model calibration, a truth case is generated by choosing h=10.4 ft $\in$ [0, 60], k_{high}=131.7 md $\in$ [100, 200] and k_{low}=1.31 md $\in$ [0, 50]. The production rates of the truth model with some added Gaussian noise represent the *known* observed data. We assume that the truth model parameters are *unknown* and that we only know the prior range of the parameters.

The aim is to update our prior knowledge about the reservoir from a set of non-informative (uniform) prior probabilities assigned to the model parameters. The McMC process requires an extensive sampling from the priors and examines how well the production rates fit the observations using the following misfit definition

$$M = 0.5 \sum_{p \in \{o,w\}} \sum_{t=1}^{T} \frac{(q_p^{obs} - q_p^{sim})_t^2}{\sigma_t^2} \qquad\qquad 2$$

in which, T is the number of observations, q_p represents oil and water rates, sim and obs indicate the simulation and observed data respectively, and σ=0.03×q_p^{obs} represents the measurement error. With the assumption of Gaussian distribution, the misfit function represents the negative log likelihood function which is used in the Bayes' rule. The lower the misfit value, the higher probability assigned to the corresponding model parameters.

To obtain the posterior distributions we need to evaluate a large number of model misfits which is not computationally feasible. Hence, to accelerate the McMC sampling process an efficient and accurate (enough) surrogate model is built. In this paper, we use the ordinary Kriging which is dynamically tuned when new samples are iteratively added by an adaptive sampling scheme. The tuning is referred to as selecting the best hyper-parameters (i.e. the covariance's length scales) that control how well the surrogate model can fit the data. This is usually involved with some optimization routines such as maximum likelihood estimation. For this problem, an initial population with 50 models is generated using a space filling Latin-Hypercube design and the black oil flow simulations are performed using tNavigator (Rock Flow Dynamics, 2016) to obtain the misfit values. tNavigator is designed to run parallel acceleration algorithms on multicore and manycore shared and distributed memory computing systems. The LOLA-Voronoi is then used to sequentially improve the quality of the selected surrogate. The LOLA-Voronoi can help construct a global surrogate by sampling from important areas of the parameter space to balance between the exploration and exploitation. The exploitation phase is used to make samples from the non-linear locations of the misfit function. This is achieved by constructing the gradients of the objective function using the nearby points (Crombecq et al., 2009). Therefore, for the places that we have a steeper slope (i.e. higher uncertainty), there are higher chances of nonlinearities which are needed to be focused. This algorithm needs to be coupled by the Voronoi-tessellation-based exploration component which is used to rank the samples based on the Voronoi cell size which is indicative of whether this region is under-sampled. The Voronoi part of the algorithm is similar to the Neighborhood Algorithm proposed by Sambridge (1999) which has already been used in many reservoir engineering applications. Because the IC fault model's misfit surface is a complicated surface with many steep local minima, the Expected Improvement statistical criterion is combined with LOLA-Voronoi to enrich the sampling process in the areas that global minima reside. After each iteration, a new sample is proposed and the flow simulation is invoked to calculate the new misfit value and to update the surrogate model with the augmented data set. Our sampling budget is set to 400, which is the maximum number of flow simulations that we can run, but also affordable. The final surrogate model is cheap-to-evaluate, which can represent the likelihood function for the extensive sampling required in the McMC process. Figure 1 (*left*) compares the efficiency of the adaptive sampling method in constructing a tuned kriging model from 400 samples to represent the IC fault model's misfit function. The figure portrays the misfit value versus the fault throw for the two fixed parameters K_{high} and K_{low} that are fixed to their corresponding truth values. This figure shows that using a one-shot Latin-Hypercube design (with 400 samples), we are unable to capture the complexity of the truth misfit surface particularly in the regions close to the steep global minimum. Figure 1 (*right*) shows the modelled misfit surface using the ordinary Kriging fitted to the samples from adaptive sampling

procedure. The black circles (i.e. the samples) are the samples proposed by LOLA-Voronoi and Expected Improvement, which are distributed non-uniformly to capture the areas with larger non-linearities. In this plot (Figure 1; *right*), K_{low} is fixed to its truth value of 1.3 md.
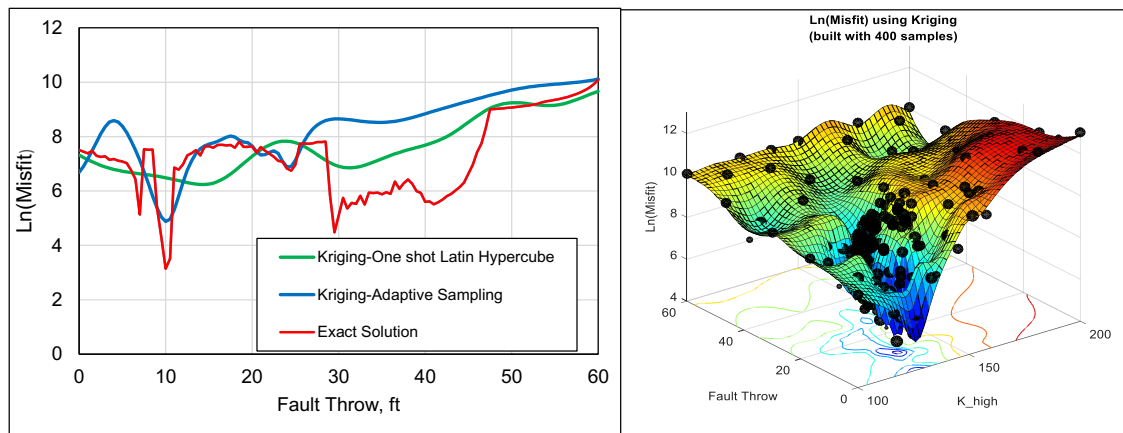

Figure 1 *Left: The accuracy and the efficiency of the adaptive sampling method to represent the IC fault's misfit surface by a Kriging proxy model as a function of fault throw. Right: The complexity of the modelled misfit function using Kriging. More samples are placed in the interested areas where the global minimum resides.*

The final tuned kriging model is used to approximate the posterior distributions. In this paper we use, a multi-chain Differential Evolution Adaptive Metropolis method (Vrugt et al., 2009) which is proved to have a superior performance compared to the classical and other single chain adaptive Metropolis methods (Vrugt, 2016). The McMC algorithm is initiated with 7 chains and this is repeated for 100000 generations. The acceptance rate is around 20% and convergence is achieved quickly. The convergence is assessed by R-statistics of Gelman and Rubin (1992) (when $\hat{R}<1.2$). Figure 2 shows the trace plots (*left*) and the marginal distributions of the posterior samples (*right*). The chains display a good mixing behavior while the modes of the distributions are located close to the truth values. The posterior samples can be eventually used to run the forward numerical simulations by a Monte Carlo sampling to provide the forecast statistics (P90-P10-P50) of the future production.
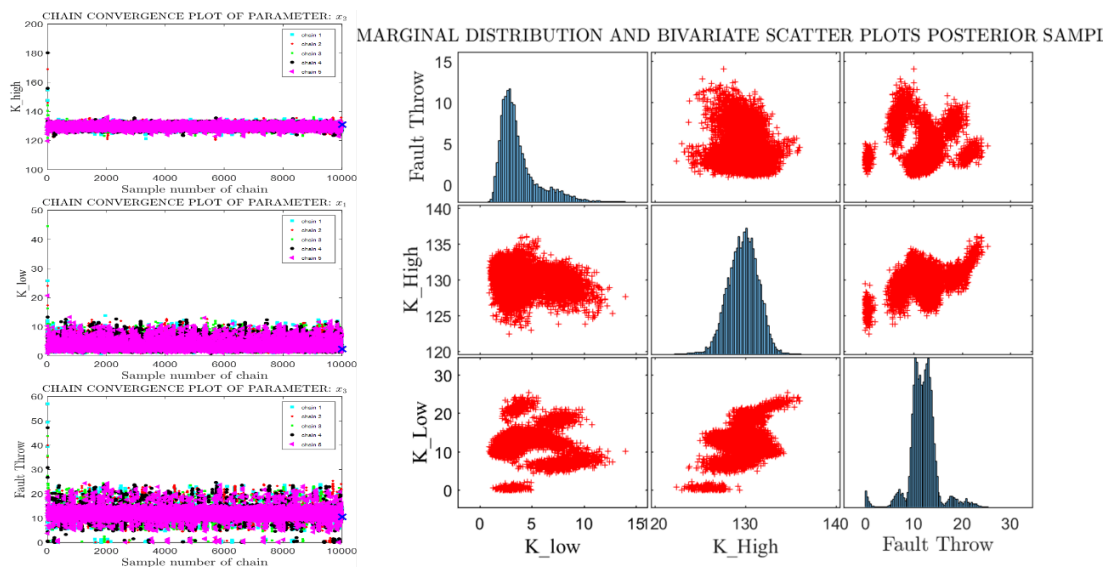

Figure 2 *The trace plots of the sampled chain trajectories (left) and the marginal posterior distributions of the IC fault model obtained from a proxy-based McMC process (right).*

Conclusions

In this paper, we presented a new method for accelerating the McMC process using a proxy-based method that is tuned by an adaptive sampling scheme (i.e. LOLA -Voronoi and Expected Improvement). The adaptive sampling method can sequentially update the Kriging proxy to create an accurate model by observing the simulation model outputs (i.e. misfits) to increase the sample density in the nonlinear regions. We applied this method to the IC fault model that has a very complicated misfit surface. The Kriging proxy was constructed using 400 samples that showed a large improvement comparing to the one-shot Latin-Hypercube design. The proxy model was successfully used to efficiently perform the McMC to find the posterior distributions. This paper is towards defining an *efficient* framework for model calibration and uncertainty quantification to reduce the computational cost.

Acknowledgements

References

Christie, M., Demyanov, V., and Erbas, D., 2006, Uncertainty quantification for porous media flows: Journal of Computational Physics, **217**(1), p. 143-158. http://dx.doi.org/http://dx.doi.org/10.1016/j.jcp.2006.01.026.

Crombecq, K., Tommasi, L.D., Gorissen, D., and Dhaene, T., 2009, A novel sequential design strategy for global surrogate modeling, *Proceedings of the 2009 Winter Simulation Conference (WSC)*, p. 731-742. http://dx.doi.org/10.1109/WSC.2009.5429687.

Gelman, A., and Rubin, D.B., 1992, Inference from Iterative Simulation Using Multiple Sequences: Statistical Science, **7**(4), p. 457-472

Jones, D., 2001, A Taxonomy of Global Optimization Methods Based on Response Surfaces: Journal of Global Optimization, **21**(4), p. 345-383. http://dx.doi.org/10.1023/A:1012771025575.

Mariethoz, P.G., and Caers, P.J., 2014, *Multiple-point Geostatistics: Stochastic Modeling with Training Images*, Wiley, 376 p

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E., 1953, Equation of State Calculations by Fast Computing Machines: The Journal of Chemical Physics, **21**(6), p. 1087-1092. http://dx.doi.org/doi: 10.1063/1.1699114.

Privault, N., 2013, *Understanding Markov Chains: Examples and Applications*, Springer Singapore

Rock Flow Dynamics, 2016, tNavigator Reservoir Simulator's user manual (v.4.2.4),

Sambridge, M., 1999, Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space: Geophysical Journal International, **138**(2), p. 479-494. http://dx.doi.org/10.1046/j.1365-246X.1999.00876.x.

Santner, T.J., Williams, B.J., and Notz, W.I., 2014, *The Design and Analysis of Computer Experiments*, Springer, 300 p

Tavassoli, Z., Carter, J.N., and King, P.R., 2005, An analysis of history matching errors: Computational Geosciences, **9**(2), p. 99-123. http://dx.doi.org/10.1007/s10596-005-9001-7.

van der Herten, J., Couckuyt, I., Deschrijver, D., and Dhaene, T., 2015, A Fuzzy Hybrid Sequential Design Strategy for Global Surrogate Modeling of High-Dimensional Computer Experiments: SIAM Journal on Scientific Computing, **37**(2), p. A1020-A1039. http://dx.doi.org/10.1137/140962437.

Vrugt, J.A., 2016, Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation: Environmental Modelling & Software, **75**, p. 273-316. http://dx.doi.org/http://dx.doi.org/10.1016/j.envsoft.2015.08.013.

Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., and Higdon, D., 2009, Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling: nternational Journal of Nonlinear Sciences and Numerical Simulation, **10**(3), p. 273-290

Yang, C., Nghiem, L., Erdle, J., Moinfar, A., Fedutenko, E., Li, H., Mirzabozorg, A., and Card, C., 2015, An Efficient and Practical Workflow for Probabilistic Forecasting of Brown Fields Constrained by Historical Data, paper SPE 175122-MS, presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas, USA. http://dx.doi.org/10.2118/175122-MS.