Adaptive multiscale MCMC algorithm for uncertainty quantification in seismic parameter estimation

Xiaosi Tan*, Texas A&M Univ.; Yalchin Efendiev, Texas A&M Univ and KAUST.; Richard L. Gibson, Jr., Texas A&M Univ.; Wing Tat Leung, Texas A&M Univ.

SUMMARY

Formulating an inverse problem in a Bayesian framework has several major advantages (Sen and Stoffa, 1996). It allows finding multiple solutions subject to flexible a priori information and performing uncertainty quantification in the inverse problem. In this paper, we consider Bayesian inversion for the parameter estimation in seismic wave propagation. The Bayes' theorem allows writing the posterior distribution via the likelihood function and the prior distribution where the latter represents our prior knowledge about physical properties. One of the popular algorithms for sampling this posterior distribution is Markov chain Monte Carlo (MCMC), which involves making proposals and calculating their acceptance probabilities. However, for large-scale problems, MCMC is prohibitively expensive as it requires many forward runs. In this paper, we propose a multilevel MCMC algorithm that employs multilevel forward simulations. Multilevel forward simulations are derived using Generalized Multiscale Finite Element Methods that we have proposed earlier (Efendiev et al., 2013a; Chung et al., 2013). Our overall Bayesian inversion approach provides a substantial speed-up *both* in the process of the sampling via preconditioning using approximate posteriors *and* the computation of the forward problems for different proposals by using the adaptive nature of multiscale methods. These aspects of the method are discussed in the paper. This paper is motivated by earlier work of M. Sen and his collaborators (Hong and Sen, 2007; Hong, 2008) who proposed the development of efficient MCMC techniques for seismic applications. In the paper, we present some preliminary numerical results.

INTRODUCTION

An important problem in seismic exploration is determining physical properties of Earth's interior from surface data and measurements. In many previous findings, the inverse problem is setup in a deterministic fashion and involves minimizing a complex functional. The deterministic approaches often do not take into account uncertainties in the data and prior information in a statistical view. A general approach that is becoming popular is the use of Bayesian inversion. In Bayesian inversion, the inverse problem is setup stochastically and the objective is to sample the posterior distribution (see Mosegaard and Tarantola, 1995, for general discussions on Bayesian inversion).

There are various methods in the literature for sampling the posterior distribution. One of the commonly used methods is the Markov chain Monte Carlo method (MCMC). This approach does not require the knowledge of the normalizing constant of the posterior distribution. The main ingredients of MCMC are (1) to make a proposal and (2) to compute the acceptance probability for the proposal. By computing the

acceptance probabilities appropriately and by a proper choice of the instrumental proposal distribution, MCMC results in a Markov chain with a steady state distribution being the desired posterior distribution, thus, it guarantees a rigorous sampling.

One of the main disadvantages of MCMC when applied to large-scale problems is the cost of computing the likelihood for each proposal (Hong and Sen, 2007; Efendiev et al., 2006, 2005). Indeed, if many proposals are made and each proposal requires solving a large-scale forward problem, this can be prohibitively expensive. In this paper, we propose a multilevel MCMC approach that use multilevel approximation of the solution efficiently. The multilevel MCMC approach consists of sampling a hierarchy of posterior distributions that satisfy certain requirements (Efendiev et al., 2013b). For constructing the hierarchical posterior distributions, we use of GMsFEM approach (Chung et al., 2013; Efendiev et al., 2013a) as discussed below.

The main idea of GMsFEM (Efendiev et al., 2013a; Chung et al., 2013) is to construct multiscale basis functions on a coarse computational grid for approximating the solution of the wave equation. The basis functions are computed using local spectral problems and selecting the dominant eigenmodes. By using only a few degrees of freedom, we can achieve a reduced-order approximation of the solution without sacrificing the accuracy. The multiscale solution strategy provides an excellent tool for constructing hierarchical approximations of the posterior distribution. In particular, the sampling from the distribution corresponding to the lowest degrees of freedom is very inexpensive with a gradual increase in the CPU as we use more basis functions in each coarse grid.

The main idea of the multilevel MCMC approach (Efendiev et al., 2013b) consists of screening the proposals based on inexpensive forward simulations. This allows conditioning the quantities of interest at one level (e.g., at a finer level) to that at another level (e.g., at a coarser level). The GMsFEM provides the mapping between the levels. More precisely, for each proposal, we run the simulations at different levels to screen the proposal and accept it conditionally at these levels. In this manner, we obtain samples from hierarchical posteriors corresponding to our multilevel approximations which can be used for rapid computations within a MLMC framework.

A second advantage of using the multiscale approach is that adaptive calculations that are inherent in multiscale basis construction. For each new proposal, one needs to update the properties only locally in the regions where the physical properties of the media has changed. Thus, only a few basis functions need to be updated. Consequently, from one sample to another sample, we can gain a substantial computational speed-up.

METHOD

Let Ω be a bounded domain. Our aim is to develop a new multiscale inverse method for the following wave equation for pressure u

$$\frac{\partial^2 u}{\partial t^2} = \nabla \cdot (a \nabla u) + f \quad \text{in } [0, T] \times \Omega \quad (1)$$

with the homogeneous Dirichlet boundary condition $u = 0$ on $[0, T] \times \partial\Omega$. The parameter a is c^2 , where c is wave speed and can be expressed as $c^2 = \kappa/\rho$, where κ is bulk modulus and ρ is density. This formulation assumes κ is constant.

Our main objective is to sample velocity field c conditioned on the observed data F_{obs} . In our numerical examples, the observed data are dynamic pressure responses at receivers that are placed at the surface. F_{obs} contains measurement errors when field data are considered. For a given velocity field c , the computed data $F(c)$ also contains modeling errors. By treating the combined error as a random variable ε we can write the model as

$$F_{obs} = F(c) + \varepsilon. \quad (2)$$

For simplicity, the noise ε will be assumed to follow a normal distribution $\mathcal{N}(0, \sigma_f^2 I)$, i.e., the likelihood $p(F_{obs}|c)$ is assumed be $\mathcal{N}(F_c, \sigma_f^2 I)$.

In our numerical example, we will take the prior field to consist of homogeneous layers separated by an unknown interface. The inverse problem consists of determining the shape of the interface, and we assume velocities known. The method could be applied to include estimation of general heterogeneous velocities and other more general parameterizations of the model. To represent the velocity field c , we let the vector τ parameterize the interface. Hence the unknown part of the velocity field is completely determined by τ . To generate a velocity field c consistent with the observed data F_{obs} , we can use Bayes' formula which expresses the posterior distribution $\pi(c)$ as

$$\pi(c) = p(c|F_{obs}) \propto p(F_{obs}|c)p(c) = p(F_{obs}|c)p(\tau),$$

In this expression for $\pi(c)$, $p(F_{obs}|c)$ is the likelihood function, incorporating the information in the data F_{obs} , and $p(\tau)$ is the prior for the parameters τ .

Sampling from the posterior distribution $\pi(c)$ conditions the velocity fields to the pressure data F_{obs} with measurement errors. This sampling can be achieved by Markov chain Monte Carlo (MCMC) methods. The main computational effort of MCMC is in evaluating the target distribution $\pi(c)$ in computing the acceptance probability (see the algorithm description in the next section).

MCMC

The standard Metropolis-Hastings algorithm (Robert and Casella, 1999) generates samples from the posterior distribution $\pi(c) = p(c|F_{obs})$, cf. Algorithm 1. Here $\mathcal{U}(0, 1)$ is the uniform distribution over the interval $(0, 1)$. Given the current sample c^m , parameterized by its parameter τ^m , one can generate the proposal

c by generating the proposal for τ first, i.e., draw τ from distribution $q_\tau(\tau|\tau^m)$, for some proposal distributions $q_\tau(\tau|\tau^m)$.

Algorithm 1 Metropolis-Hastings MCMC

```

1: Specify  $c_0$  and  $M$ .
2: for  $m = 0 : M$  do
3:   Generate the velocity field proposal  $c$  from  $q(c|c^m)$ .
4:   Compute the acceptance probability  $\gamma(c^m)$  (cf. (3), see
      (Robert and Casella, 1999) for more details)
5:   Draw  $u \sim \mathcal{U}(0, 1)$ .
6:   if  $\gamma(c^m, c) \leq u$  then
7:      $c^{m+1} = c$ .
8:   else
9:      $c^{m+1} = c^m$ .
10:  end if
11: end for

```

The main disadvantage of MCMC algorithm is the high computational cost in the forward simulations to compute the likelihood in the target distribution $\pi(c)$. Typically, MCMC method in simulations converges to the steady state very slow and the acceptance rate is also very low. A large amount of CPU time is spent on simulating the rejected samples. For this reason, we propose multilevel MCMC.

MULTILEVEL MCMC

We introduce a multilevel MCMC algorithm by adapting the proposal distribution $q(c|c_m)$ to the target distribution $\pi(c)$. We will use hierarchical posteriors using the GMsFEM with different sizes of the basis space which we call different levels, cf. Algorithm 2. The process modifies the instrumental proposal distribution $q(c|c_m)$. We denote by $F_l(c)$ the data computed by solving coarse problem at level l for a given c . The target distribution $\pi(c)$ is approximated on level l by $\pi_l(c)$, with $\pi(c) \equiv \pi_L(c)$. Here we have

$$\pi_l(c) \propto \exp\left(-\frac{\|F_{obs} - F_l(c)\|^2}{2\sigma_l^2}\right) \times p(c). \quad (6)$$

Below, we describe multilevel MCMC algorithm that uses hierarchical posterior distribution. We have investigated the convergence of this algorithm in Efendiev et al. (2013b).

HIERARCHICAL AND ADAPTIVE MODELING WITH GMSFEM

In this section, we briefly describe the multiscale method, discontinuous Galerkin GMsFEM, that is used to perform adaptive and hierarchical simulations. The main point of the GMsFEM is that we have to construct a small number of multiscale basis functions that capture the effects of small scale features reliably.

We use the Interior Penalty Discontinuous Galerkin (IPDG) method (Grote and Schotzau, 2009; Basabe et al., 2008) to

Multiscale MCMC

Algorithm 2 Multilevel Metropolis-Hastings MCMC

- 1: Given c_m , make a proposal c from instrumental distribution $q(c|c_1^m) = q_0(c|c_1^m)$
- 2: Compute the acceptance probability

$$\rho_1(c_1^m, c) = \min \left\{ 1, \frac{q_0(c_1^m|c)\pi_1(c)}{q_0(c|c_1^m)\pi_1(c_1^m)} \right\} \quad (3)$$

- 3: Accept c with probability $\rho_1(c_1^m, c)$ and set $c_1^{m+1} = c$ if accepted, or $c_1^{m+1} = c_1^m$, otherwise
- 4: **for** $l = 1 : L - 1$ **do**
- 5: **if** c is accepted at level l **then**
- 6: Form the proposal distribution q_l (on the $l + 1$ th level) by

$$q_l(c|c_{l+1}^m) = \rho_l(c_{l+1}^m, k)q_{l-1}(c|c_{l+1}^m) + \delta_{c_{l+1}^m} (1 - \int \rho_l(c_{l+1}^m, k)q_{l-1}(c|c_{l+1}^m)dc_{l+1}^m) \quad (4)$$

- 7: Compute the acceptance probability

$$\rho_{l+1}(c_{l+1}^m, c) = \min \left\{ 1, \frac{q_l(c_{l+1}^m|c)\pi_{l+1}(c)}{q_l(c|c_{l+1}^m)\pi_{l+1}(c_{l+1}^m)} \right\} = \min \left\{ 1, \frac{\pi_l(c_{l+1}^m)\pi_{l+1}(c)}{\pi_l(c)\pi_{l+1}(c_{l+1}^m)} \right\} \quad (5)$$

- 8: Accept c with the probability $\rho_{l+1}(c_{l+1}^m, c)$, and set $c_{l+1}^{m+1} = c$, if accepted, or $c_{l+1}^{m+1} = c_{l+1}^m$, otherwise
- 9: **end if**
- 10: **end for**

couple multiscale basis functions. For each interior coarse grid edge e , we let K^- and K^+ be the two coarse grid blocks having the common coarse edge e . Then we define the average and the jump operators respectively by $\{v\}_e = \frac{v^+ + v^-}{2}$, $[u]_e = u^+ - u^-$, where $u^\pm = u|_{K^\pm}$ and we have assumed that the normal vector on e is pointing from K^+ to K^- . For each coarse edge e that lies on the boundary of Ω , we define $\{v\}_e = v$ and $[u]_e = u$ assuming the unit normal vector on e is pointing outside the domain. Let $V_H = \{\phi_i\}_{i=1}^N$ be a finite dimensional function space which consists of functions $\phi_i(x)$ that are smooth in each coarse grid block but are in general discontinuous across coarse grid edges. Let u_H be the numerical solution obtained by our GMsFEM. Then u_H can be represented by

$$u_H(x, z, t) = \sum_{i=1}^N d_i(t) \phi_i(x). \quad (7)$$

The IPDG method can then be stated as: find u_H such that

$$\int_{\Omega} \frac{\partial u_H}{\partial t^2} v + a_{DG}(u_H, v) = \int_{\Omega} f v, \quad \forall v \in V_H, \quad (8)$$

where the bilinear form $a_{DG}(u, v)$ is defined by

$$a_{DG}(u, v) = \sum_{K \in \mathcal{T}_H} \int_K a \nabla u \cdot \nabla v + \sum_{e \in \mathcal{E}_H} \left(- \int_e \{a \nabla u \cdot n\}_e [v]_e - \int_e \{a \nabla v \cdot n\}_e [u]_e + \frac{\gamma}{h} \int_e a [u]_e [v]_e \right)$$

where $\gamma > 0$ is a penalty parameter and n denotes the unit normal vector on e . We define the mass matrix M , stiffness matrix K and the right hand side vector F by

$$M_{ij} = \int_{\Omega} \phi_i \phi_j, \quad K_{ij} = a_{DG}(\phi_i, \phi_j), \quad F_i = \int_{\Omega} f \phi_i. \quad (9)$$

Then equation (8) can be written as

$$M \frac{d^2 U}{dt^2} + KU = F \quad (10)$$

where U is a vector defined by $U = (d_i(t))$. Let $\Delta t > 0$ be the time step size and let $U^n = (d_i(t_n))$. The time discretization is performed by the classical second order central finite difference method; we find U^{n+1} such that

$$M \frac{U^{n+1} - 2U^n + U^{n-1}}{\Delta t^2} + KU^n = F^n.$$

Next, we will discuss the choice of V_H . For each coarse grid block K , we let $V_H(K)$ be the restriction of V_H on K . The space $V_H(K)$ is divided into two orthogonal components, namely, $V_H(K) = V_H^1(K) + V_H^2(K)$. The first component $V_H^1(K)$ takes care the contribution of the solution on coarse grid boundaries. In particular, for a given choice of boundary condition on coarse grid boundary, we extend it to the interior by solving

$$\nabla \cdot (a \nabla w_{i,K}) = 0 \quad (11)$$

with the prescribed boundary condition. In our numerical simulations, we do not need to use all of these basis functions and use a local spectral problem to identify multiscale basis functions by choosing dominant modes. Given the functions $w_{i,K}$, we define two matrices A and B by

$$A_{ij} = \frac{1}{h} \int_{\partial K} w_{i,K} w_{j,K}, \quad (12)$$

$$B_{ij} = \int_K a \nabla w_{i,K} \cdot \nabla w_{j,K} + \frac{1}{h} \int_{\partial K} w_{i,K} w_{j,K}.$$

Then we consider the spectral problem

$$Aq_i = \mu_i Bq_i. \quad (13)$$

The eigenvalues are ordered in decreasing order, namely, $\mu_1 \geq \mu_2 \geq \dots$ and the corresponding eigenfunctions are denoted by q_i . The most dominant mode is represented by the eigenfunctions with large eigenvalue. We select the first few, say L , eigenfunctions so that the sum of the corresponding eigenvalues is a large percentage of the total energy E , which is the sum of all eigenvalues. The space $V_H^1(K)$ is then spanned by the functions $\sum (q_i)_j w_{j,K}$, $i = 1, 2, \dots, L$.

The space $V_H^2(K)$ contains the standing modes that are zero on the boundary of the coarse grid K by considering the following eigenvalue problem $-\nabla \cdot (a \nabla z_\lambda) = \lambda z_\lambda$. As before, we order the eigenvalues and choose dominant modes to construct multiscale basis functions.

Finally, we note that GMsFEM can be used to do adaptive simulations. For each new proposal, the multiscale basis functions only near the interface are updated while keeping the basis functions away from the interface unchanged. This provides an additional substantial speed-up in forward simulations.

Multiscale MCMC

Figure 1: Plot of error versus iteration.

NUMERICAL RESULTS

In our numerical experiment, we consider a 1000 m by 1000 m model consisting of two homogeneous layers. The upper layer has wave speed $c_1 = 2500$ m/s, density $\rho_1 = 2600$ kg/m³, and lower layer has wave speed $c_2 = 2800$ m/s and density $\rho_2 = 2700$ kg/m³. The source term is the Ricker wavelet,

$$f(x, z, t) = g(x, z)(1 - 2\pi^2 f_0^2 (t - 2/f_0)^2) e^{-\pi^2 f_0^2 (t - 2/f_0)^2},$$

where the frequency $f_0 = 20$ Hz and $g(x, z)$ is a spatial Gaussian point source located at (500, 100) and defined by

$$g(x, z) = H^{-2} e^{-H^{-2}((x-500)^2 + (z-100)^2)}.$$

The interface of the two layers is generated by an interpolating parametric cubic spline based on 11 specific parameter points. Here we impose an apriori information that the interface is bounded between the depth 400 m to 600 m. The instrumental proposal distribution is a random walk sampler ((Robert and Casella, 1999)).

We will show the result for a two-level MCMC simulation. For GMsFEM, we take 20 boundary basis functions (in V_H^1) and 5 interior basis functions (in V_H^2) for the first level, 25 boundary basis functions (in V_H^1) and 10 interior basis functions (in V_H^2) for the second (finer) level. Our quantity of interest, the pressure field F_{obs} , is taken as the pressure measured at 20 equally distributed receiver locations near the top boundary.

The acceptance rate on the first level (pre-screening level) of multilevel MCMC is 12.6%, and on the finer level, the acceptance rate is 88.9%. For more expensive levels, we observe that it is much more probable that a proposed sample will be accepted. The plot in Figure 1 illustrates the error $E_c = \|F_{obs} - F_c\|$ of the samples on the finer grid. We can see the combined error decreases and remains low, showing that the Markov Chain has converged. Figure 2 displays the reference velocity field and the best fitted sample (with least error to reference). In Figure 3, we show the mean of the accepted velocity field on the fine level and the mean interface depth from the accepted samples with the parametric spline points. We can see that the mean interface has a shape similar to that of the reference. Finally, Figure 4 shows the seismogram from $t = 0$ to 0.5. The blue curves are the reference observation, the red ones result for the initial model, and the green ones are the result of the best fitting model.

Figure 2: Reference velocity field (left) and the sample field giving best fit to data (right). (m/s)

Figure 3: Posterior mean (left) and mean interface shape from the inversion (right). (m/s)

Figure 4: Comparison of seismograms for: Blue: reference solution; Red: initial guess; Green: Best fitted solution

CONCLUSIONS

We present a multilevel MCMC approach coupled with hierarchical and adaptive forward models that use GMsFEM to efficiently solve Bayesian inversion problems that arise in parameter estimation. In multilevel MCMC, for each proposal, we use a hierarchical posterior distributions computing using different reduced-order models on a coarse grid to screen the proposal. The latter is important for eliminating bad proposals that will be rejected after costly forward simulations. Applying GMsFEM adaptively, we update multiscale basis function around the interface and this allows avoiding costly forward simulation for new proposals. We consider an example that consists of two layered media with the objective is to determine the location of the interface and quantify uncertainties associated with the inverse problem. Our preliminary numerical results show that the proposed method is robust. We plan to apply our methods to more complex inverse problems in future.

EDITED REFERENCES

Note: This reference list is a copy-edited version of the reference list submitted by the author. Reference lists for the 2014 SEG Technical Program Expanded Abstracts have been copy edited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

- De Basabe, J. D., M. K. Sen, and M. F. Wheeler, 2008, The interior penalty discontinuous Galerkin method for elastic wave propagation: Grid dispersion: *Geophysical Journal International*, **175**, no. 1, 83–93, <http://dx.doi.org/10.1111/j.1365-246X.2008.03915.x>.
- Chung, E., W. Leung, Y. Efendiev, and R. Gibson Jr., 2013, Generalized multiscale finite element modeling of acoustic wave propagation: Presented at the 83rd Annual International Meeting, SEG.
- Efendiev, Y., A. Datta-Gupta, V. Ginting, X. Ma, and B. Mallik, 2005, An efficient two-stage Markov chain Monte Carlo method for dynamic data integration: *WRR*, **41**, no. 12, W1242, <http://dx.doi.org/10.1029/2004WR003764>.
- Efendiev, Y., J. Galvis, and T. Y. Hou, 2013a, Generalized multiscale finite element methods (GMsFEM): *Journal of Computational Physics*, **251**, 116–135, <http://dx.doi.org/10.1016/j.jcp.2013.04.045>.
- Efendiev, Y., T. Hou, and W. Luo, 2006, Preconditioning Markov chain Monte Carlo simulations using coarse-scale models: *SIAM*, **28**, no. 2, 776–803, <http://dx.doi.org/10.1137/050628568>.
- Grote, M., and D. Schotzau, 2009, Optimal error estimates for the fully discrete interior penalty DG method for the wave equation: *Journal of Scientific Computing*, **40**, no. 1–3, 257–272, <http://dx.doi.org/10.1007/s10915-008-9247-z>.
- Hong, T., 2008, Mcmc algorithm, integrated 4D seismic reservoir characterization and uncertainty analysis in a Bayesian framework: Ph.D. thesis, University of Texas.
- Hong, T., and M. K. Sen, 2007, A new MCMC algorithm for seismic parameter estimation and uncertainty analysis: 77th Annual International Meeting, SEG, Expanded Abstracts, **26**, 1883–1887.
- Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: *Journal of Geophysical Research*, **12**, 431–447.
- Robert, C., and G. Casella, 1999, Monte Carlo statistical methods: Springer-Verlag.
- Sen, M., and P. L. Stoffa, 1996, Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion: *Geophysical Prospecting*, **44**, no. 2, 313–350, <http://dx.doi.org/10.1111/j.1365-2478.1996.tb00152.x>.