Bayesian Inversion with an Improved Proposal Distribution

Reza Ghanati[1] Zahra Tafaghod Khabaz[2]
[1]*Assistant Professor, Institute of Geophysics, University of Tehran, rghanati@ut.ac.ir*
[2] *M.Sc. Graduated, Institute of Geophysics, University of Tehran, zahra.tafaghod@ut.ac.ir*

ABSTRACT

Geophysical inverse problems are generally formulated as an ill-posed non-linear optimization problem commonly solved through deterministic gradient-based approaches. Using these methods, despite fast convergence properties, may lead to local minima as well as impend accurate uncertainty analysis. On the contrary, formulating a geophysical inverse problem in a probabilistic framework and solving it by constructing the multi-dimensional posterior probability density (PPD) allow for complete sampling of the parameter space and uncertainty quantification. The PDD is numerically characterized using Markov Chain Monte Carlo (MCMC) approaches. However, the convergence of the MCMC algorithm (i.e. sampling efficiency) toward the target stationary distribution highly depends upon the choice of the proposal distribution. In this paper, we develop an efficient proposal distribution based on perturbing the model parameters through an eigenvalue decomposition of the model covariance matrix in a principal component space. The proposed strategy is illustrated for the inversion of synthetic and real geo-electrical data sets. The numerical experiments demonstrate that the presented proposal distribution takes advantage of the benefits from an accelerated convergence and mixing rate compared to the conventional Gaussian proposal distribution.

Keywords: MCMC, Principal component space, Proposal distribution.

INTRODUCTION

Geophysical inverse problems seek to provide quantitative information about geophysical characteristics of the Earth's subsurface for indirectly related data and measurements. Geophysical inverse problems are generally formulated as an ill-posed non-linear optimization problem commonly solved through Newton-based methods. The significant property of the gradient-based approaches is their fast convergence toward the final solution, but the local linearization of the inverse solution hinders reliable uncertainty appraisal. A remedy to the exiting limitations is to employ derivative-free global direct search techniques such as Markov Chain Monte Carlo (MCMC) algorithms in the Bayesian framework. The MCMC sampling is essentially a guided-random walk through the probable parts of the posterior model space. In recent years, several variants of the MCMC algorithms have been proposed in the mathematical and geophysical literature. Beginning from the Metropolis-Hastings algorithm and Gibbs method (Metropolis and Ulam 1949), different investigations on enhancing the efficiency of the MCMC algorithms have been implemented for better performance of the classical approaches. For instance, Haario et al (2001) proposed a self-tuning algorithm namely the adaptive metropolis algorithm. Later, they suggested the delayed rejection adaptive Metropolis method (Haario et al, 2006). Ter Braak (2006) introduced the differential evolution Monte Carlo algorithm. Vrugt et al (2008) applied the idea of combining the differential evolution technique and the adaptive Metropolis to hydrological data. An accelerated variant of the MCMC technique is the parallel tempering algorithm (Swendsen and Wang, 1986). Recent examples of parallel tempering as applied to geophysical inversion can be found in Ray et al (2013), and Blatter et al (2021). Despite significant improvements in MCMC sampling methods, choosing an appropriate proposal distribution to generate trial moves in the Markov chain, is of crucial importance for an efficient inversion process that creates a well-mixed Markov-Chain. The common choice for proposal density is a Gaussian distribution centered on the current model and variance tuned by the user. The variance parameter determines the step length at each MCMC iteration. When the width of the proposal distribution is too wide, the acceptance ratio is small, and therefore the chain will not mix efficiently and converge only slowly to the target distribution. On the other hand, if the width of the proposal distribution is too narrow, nearly all candidate models are accepted, but the chain mixes again very slowly. In this paper, to improve the functionality of the Metropolis-Hastings sampler in terms of the faster convergence of the chain and good mixing properties, we develop an efficient proposal distribution based on perturbing the model parameters through an eigenvalue decomposition of the model covariance matrix in a principal component space. The covariance matrix is retrieved from an early burn-in sampling, which itself commences using a linearized covariance estimate. The proposed algorithm is first applied for nonlinear inversion of synthetic and real geo-electrical sounding data based on multiple depth layers of the fixed boundary.

METHODOLOGY

By the rules of the conditional probabilities, the posterior can be inverted based on the Bayes' theorem in which the prior probability distribution and the likelihood function are combined and formulated as (Tarantola, 2005):

$$\Omega(\mathbf{m}|\boldsymbol{d}) = \frac{\Omega(\boldsymbol{d}|\mathbf{m})\Omega(\mathbf{m})}{\Omega(\boldsymbol{d})} \tag{1}$$

The term $\Omega(\boldsymbol{d}|\mathbf{m})$ can be interpreted as the likelihood function ($L(\boldsymbol{d}|\mathbf{m})$) which is the density function of the observed data $\boldsymbol{d} \in \mathfrak{R}^{m\times 1}$ given the model parameters $\mathbf{m} \in \mathfrak{R}^{n\times 1}$. The likelihood function depends on the statistics of the noise distribution. The unconditional distribution of the unknowns, $P(\mathbf{m})$, is called the prior distribution. This describes the knowledge about the unknowns that existed before or exists independent of the current observations. The quantity $\Omega(\boldsymbol{d})$, known as the evidence, is a normalization constant.

Assuming that the data error to be independent uncorrelated zero-mean Gaussian, we define the likelihood function by the relation (Tarantola, 2005):

$$\log L(\boldsymbol{d}|\mathbf{m}) = -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log\left(\prod_{i=1}^{m}\sigma_i^2\right) - \frac{1}{2}\big(\boldsymbol{d} - \boldsymbol{F}(\mathbf{m})\big)^T \boldsymbol{C}_d^{-1}(\boldsymbol{d} - \boldsymbol{F}(\mathbf{m})) \tag{2}$$

where $\boldsymbol{C}_d$ is the data error covariance matrix, $\boldsymbol{F}$ is the non-linear forward operator, and σ is the data error.

We incorporate smoothness constraints into the model parameters by imposing independent normal distributions to the vertical model gradient. Hence, we define zero-mean normal prior distributions concerning the vertical resistivity gradient as follows:

$$\log p(\mathbf{m}) = -\log(2\pi\beta^2) - \frac{1}{2\beta^2}(\mathbf{m}^T\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\mathbf{m}) \tag{3}$$

where β is analogous to model regularization weights used in the deterministic inversion. To effectively solve the inverse problem, the posterior distribution $\Omega(\mathbf{m}|\boldsymbol{d})$ is sampled by the Metropolis-Hastings (MH) sampler (Hastings, 1970) which is an MCMC algorithm. The MH algorithm draws a sequence of random samples from the posterior density that has the ergodic property. The MH method proceeds in two steps. At the first step, a candidate model $\boldsymbol{m}^p$ is created using a proposal density $p(\mathbf{m}^p|\mathbf{m}^c)$ based on the current model $\mathbf{m}^c$, at the second step, it is decided that the candidate model is either accepted or rejected using the MH acceptance condition. The acceptance probability, $\alpha(\mathbf{m}^p|\mathbf{m}^c)$, of the candidate model judges whether the new model is accepted to the Markov Chain or rejected. The acceptance probability can be written as

$$\alpha(\mathbf{m}^p|\mathbf{m}^c) = \min\,(1,\ \frac{\mathrm{p}(\boldsymbol{d}|\mathbf{m}^p)\mathrm{p}(\mathbf{m}^p)}{\mathrm{p}(\boldsymbol{d}|\mathbf{m}^c)\mathrm{p}(\mathbf{m}^c)}) \tag{4}$$

The usual choice for proposal density is a Gaussian distribution centered on the current model and standard deviation adjusted by the user. Using this strategy for cases with many model parameters is usually laborious in terms of computing time and user input since many short trial runs have to be made. To overcome this difficulty, we present an efficient proposal scheme based on the principal component transformation and eigenvectors decomposition of the covariance matrix estimated from the deterministic inversion. To generate new (perturbed) model parameters using our proposal method (PCA-based proposal distribution), one needs to implement the following stages:

1) Calculate the model covariance matrix using the linearized inversion (deterministic inversion) results, that is, $\boldsymbol{C}_m = J^{\dagger}J$, where J is the jacobian matrix and $J^{\dagger}$ is the regularized generalized inverse of the matrix J.
2) Decompose the covariance matrix $\boldsymbol{C}_m$ in terms of the diagonal matrix $\boldsymbol{D}$ of eigenvalues and matrix $\boldsymbol{V}$ whose columns are the corresponding right eigenvectors
3) Transform the model parameters (the current model) into principal component space leading to the rotated parameters: $\widehat{\mathbf{m}} = \boldsymbol{V}^T\mathbf{m}$, where T is the transpose.
4) Draw parameter perturbations using a Cauchy proposal distribution with $\boldsymbol{D}$ degrees of freedom (scaling factor). In the Matlab environment, $\mathbf{m}^{perturb} = \mathrm{trnd}(\mathrm{diag}(\boldsymbol{D}),m,1)$, where m is the number of model parameters.
5) Calculate the candidate model (proposal model) through adding the rotated parameters ($\widehat{\mathbf{m}}$) to the parameter perturbations derived from step 4, that is, $\widehat{\mathbf{m}} = \widehat{\mathbf{m}} + \mathbf{m}^{perturb}$
6) Rotate the proposal model back to the original parameters $\mathbf{m} = \boldsymbol{V}\widehat{\mathbf{m}}$.

Step (1) is implemented only at the first iteration. Note that the starting model of the MH algorithm corresponding to the first iteration is obtained from the linearized inversion. As earlier mentioned, the covariance matrix obtained from the deterministic inversion is used during the burn-in phase, and as sampling progresses the covariance matrix is adaptively replaced with an alternative computed based on averaging over successive models along the Markov chain, as follows:

$$\boldsymbol{W}_{\mathbf{m}} = \sum_{i=1}^{n}(\mathbf{m}_i - \mathbf{m}_{i-1})(\mathbf{m}_i - \mathbf{m}_{i-1})^T \qquad (5)$$

.NUMERICAL EXAMPLE

We provide a field data set with known geology information aiming at appraising the efficacy of the proposed procedure. We compared the results of the proposed scheme with those of the conventional method. The field data has been acquired on the German North Sea Island Borkum, utilizing the Schlumberger configuration including 23 apparent resistivity records with current electrode spacing ranging logarithmically from 1.5 m to 150 m. The geology section is illustrated by the lithology log obtained through the borehole data (see Figure 1). We implement the MCMC sampler with the Gaussian proposal density and the PCA-based proposal distribution for 300,000 and 150,000 iterations, respectively. Figures 1-top panel shows the result provided by the conventional approach whereas Figures 1-bottom panel shows the result provided by the PCA proposal density. When viewing Figures 1, we notice that both strategies give rise to relatively similar results, however, the geo-electric profiles of the Gaussian proposal density show higher variability than the PCA proposal procedure. In other words, the regions with larger uncertainty are characterized by broader distribution. Referring to the resulting geo-electric profiles and as would be expected since the variant geometry inversion requires a priori information about the layer thicknesses, the estimated uncertainty is consistently larger than the invariant geometry inversion. We also produce individual trace plots for four neighboring model parameters, ($\mathbf{m}_2, \mathbf{m}_4, \mathbf{m}_6$, and $\mathbf{m}_8$) of the subsurface layers obtained by applying a Gaussian proposal density (see Figure 2-left), and by applying the PCA proposal method (see Figure 2-right) to assess the chain convergence.

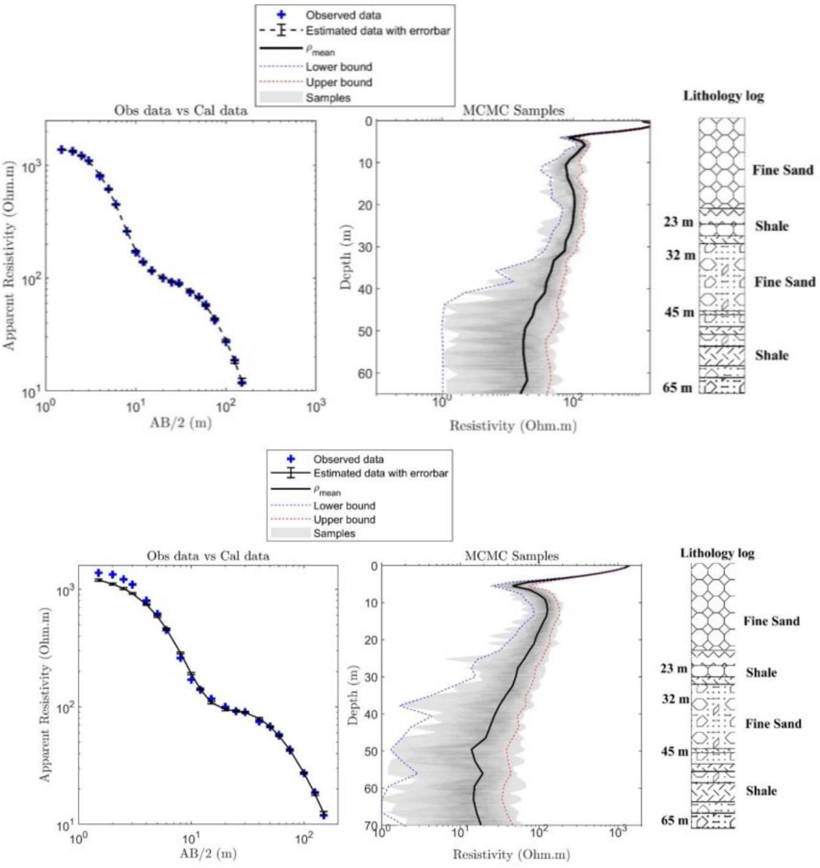

Figure 1. Right panels: Posterior model results derived from the Gaussian proposal density-based MCMC method (top-right panel) and the PCA proposal density-based MCMC method (bottom-right panel) for the real data set, Left panels: observed data versus calculated data with the error bars. The black line shows the mean posterior models and the blue line indicates true parameter values. Lower and upper intervals computed based on 95 percent confidence interval are displayed by dashed blue and red lines, respectively.

Therein, the Gaussian proposal scheme exhibits a slower mixing rate and higher correlation for the corresponding parameters leading to not extensively exploring the sample space compared to the proposed algorithm. Furthermore, our algorithm has reached the stationary region of the target density with less number of iterations (here, the trace plot of posterior samples are shown after discarding the burn-in phase samples) compared to the conventional method.

CONCLUSIONS

The Bayesian inversion success depends on the shape and size of the proposal distribution used to implement the MCMC algorithm. Choosing an appropriate proposal density for the Metropolis-Hastings acceptance step is non-trivial. As the rejection rate increases, the computational cost for accepting a proposal sample increases. To overcome this problem, we developed an adaptive Metropolis-Hastings based on perturbing the model parameters through an eigenvalue decomposition of the model covariance matrix in a principal component space. This approach applies perturbations in principal component space where rotated parameters are uncorrelated, with the eigenvectors providing the rotation matrix and the eigenvalues providing appropriate perturbation length scales. The proposed scheme was examined for inversion of real geo-electrical sounding data based on multiple depth layers of the fixed boundary. The numerical results indicated that the PCA perturbation proposal scheme significantly improved the efficiency of the MCMC sampler over the conventional Gaussian perturbation in terms of the faster convergence of the chain and good mixing properties.

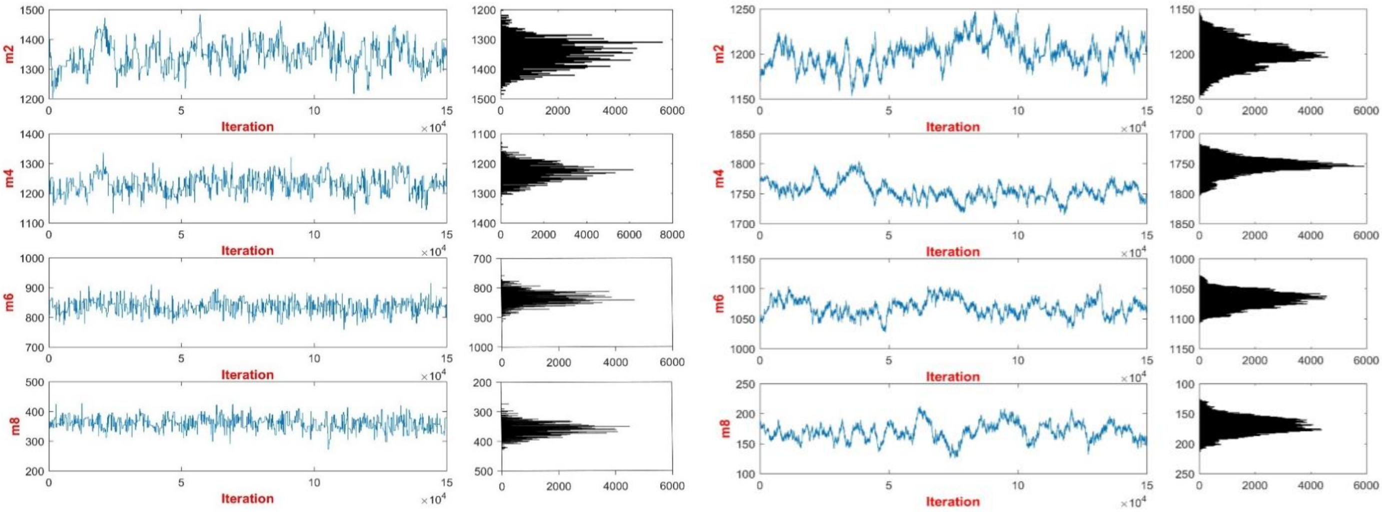

Figure 2. The trace plot of posterior samples corresponds to the real example generated by the conventional Gaussian proposal distribution-based MCMC sampling method on the left and the resulting histogram on the right and by the presented proposal distribution-based MCMC sampling method on the right and the resulting histogram. Here, only four model parameters variation in terms of different iterations is shown.

REFERENCES

Blatter, D., Ray, A., Key, K., 2021, Two-dimensional Bayesian inversion of magneto-telluric data using trans-dimensional Gaussian processes, Geophysical Journal International, 226(1) 548–563.

Haario, H., Saksman, E., and Tamminen, J., 2001, an adaptive Metropolis algorithm, Bernoulli, 7, 223-242.

Haario, H., Laine, M., Mira, A., and Saksman, E., 2006, DRAM: Efficient adaptive MCMC, Statistics and Computing, 16, 339-354.

Hastings, H., 1970, Monte Carlo sampling methods using Markov chains and their applications, Biometrika, 57, 97-109.

Metropolis, N., and Ulam, S., 1949. The Monte Carlo method, Journal of American Statistics Association, 44, 335–341.

Ray, A., Alumbaugh, D. L., Hoversten, G. M. and Key, K., 2013, Robust and accelerated Bayesian inversion of Marine controlled-source electromagnetic data using parallel tempering, Geophysics, 78(6), E271–E280.

Swendsen R. H., and Wang, J. S., 1987, Non-universal Critical dynamics in Monte Carlo simulations, Physics Review Letter 58, 86–88.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation. SIAM, 342.

Ter Braak, C. J. F., 2006, A Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces, Statistics and Computing, 16, 239-249.

Vrugt, J. A., Ter Braak, C. F. F., Gupta, H. V., and Robinson, B. A., 2008, Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling. Stochastic Environmental Research and Risk Assessment, 23(7), 1011–1026.