Fast uncertainty quantification for 2D full-waveform inversion with randomized source subsampling

Zhilong Fang, Felix J. Herrmann and Curt D. Silva
Department of Earth and Ocean sciences, University of British Columbia, Vancouver, BC, Canada

January 17, 2014

Abstract

Uncertainties arise in every area of seismic exploration, especially in full-waveform inversion, which is highly non-linear. In the framework of Bayesian inference, uncertainties can be analyzed by sampling the posterior probability density distribution with a Markov chain Monte-Carlo (McMC) method. We reduce the cost of computing the posterior distribution by working with randomized subsets of sources. These approximations, together with the Gaussian assumption and approximation of the Hessian, leads to a computational tractable uncertainty quantification. Application of this approach to a synthetic leads to standard deviations and confidence intervals that are qualitatively consistent with our expectations.

Introduction

Uncertainties relating to acquisition noise, numerical errors, and model errors, arise in every area of seismic exploration such as acquisition, velocity analysis, migration and full-waveform inversion (FWI). Because of this, quantifying the uncertainty of inversion results has become increasingly important in order to inform decisions in industry. However, due to the high computational cost associated with solving large numbers of wave equations, there has been little development on uncertainty quantification (UQ) in the context of full-waveform inversion.

In the framework of Bayesian inference, uncertainties are analyzed by sampling the posterior probability distribution. Amongst the relatively few research findings in uncertainty analysis for FWI, the Markov chain Monte Carlo (McMC) method involving local Hessian information of the negative log posterior (Martin et al., 2012) is one of the most efficient methods, because it allows for sampling the posterior distribution efficiently with the information of the Hessian. While this approach is computationally efficient, several challenges remain, namely (i) how to efficiently calculate a solution to FWI that corresponds to the most likely estimate or the maximum a posteriori (MAP) estimate of the posterior distribution; (ii) approximating the Hessian involved in calculating the uncertainty efficiently; and (iii) calculating the posterior probability function (PDF) of realizations efficiently. Addressing these issues in the context of FWI is extremely challenging because of the cost of evaluating the objective, gradient, and the realizations from the posterior distribution, all of which involve a large numbers of PDE solves (Bui-Thanh et al., 2013). Our main contribution in this work is that we reduce the computational costs of the McMC method (Martin et al., 2012; Bui-Thanh et al., 2013) significantly by evaluating the posterior PDF over randomly selected subsets of shots.

Methodology

Bayesian statistics offers a natural framework for uncertainty quantification of the subsurface velocity model, because it allows us to easily incorporate information from observations, forward maps, and prior knowledge to analyze the velocity model. According to the Bayesian theory, the posterior distribution $\pi_{\text{post}}(\mathbf{m})$ of velocity model $\mathbf{m}$ is understood as a conditional probability $\pi(\mathbf{m}|\mathbf{d}_{\text{obs}})$ of the model given the observed data $\mathbf{d}_{\text{obs}}$, which is proportional to the product of the prior on the model $\pi_{\text{prior}}(\mathbf{m})$ and the probability of the observed data given the model $\pi(\mathbf{d}_{\text{obs}}|\mathbf{m})$ — i.e., we have

$$\pi_{\text{post}}(\mathbf{m}) := \pi(\mathbf{m}|\mathbf{d}_{\text{obs}}) \propto \pi_{\text{prior}}(\mathbf{m})\pi(\mathbf{d}_{\text{obs}}|\mathbf{m}). \tag{1}$$

By sampling the posterior distribution, we can obtain information on the uncertainties and estimate statistical properties of the velocity model that include standard deviation, confidence interval, and covariance.

If we assume the distributions for the noise and the prior for velocity model to be Gaussian, the posterior PDF of 2D acoustic full-waveform inversion can be written explicitly (within a normalizing constant) as

$$\pi_{\text{post}}(\mathbf{m}) \propto \exp[-\frac{1}{2}\|f(\mathbf{m}) - \mathbf{d}_{\text{obs}}\|^2_{\Gamma^{-1}_{\text{noise}}} - \frac{1}{2}\|\mathbf{m} - \overline{\mathbf{m}}_{\text{prior}}\|^2_{\Gamma^{-1}_{\text{prior}}}], \tag{2}$$

where $f(\mathbf{m})$ represents the forward map, which generates predicted data of N_s shots for a given velocity model. Because the PDFs are assumed to be Gaussian, the distribution is completely characterized by the mean and covariance, i.e., we have $\mathcal{N}(0,\Gamma_{\text{noise}})$ for the noise and $\mathcal{N}(\overline{\mathbf{m}}_{\text{prior}},\Gamma_{\text{prior}})$ for the model.

Uncertainty analysis for FWI based on Eq. (2) is challenging for following reasons. Firstly, the parameter-to-observable map $f(\mathbf{m})$ is nonlinear and computing $f(\mathbf{m})$ is expensive because it involves solving wave equations for many shots and frequencies. Secondly, there are a large number of model parameters, especially at high frequencies, which further increases the cost of evaluating $f(\mathbf{m})$. Finally, the McMC method applied to FWI relies on generating random samples that requires computation of the inverse of the Hessian, which is also very expensive. To overcome these computational challenges, we approximate the data misfit term in Eq. (2) by randomized source subsampling. This results in a fast McMC method that consists of the following six steps, which are similar to Bui-Thanh et al. (2013):

1. *MAP estimate*: To generate samples for the velocity model, we need to center the Gaussian distribution. Following Bui-Thanh et al. (2013), we use the MAP estimate for this purpose and this involves solving the conventional FWI problem for all shots. Alternatively, we can calculate the MAP estimate using computationally efficient formulations of FWI such as the stochastic optimization method (van Leeuwen and Herrmann, 2013b) and the frugal method recently proposed by van Leeuwen and Herrmann (2013a). These methods can easily lead to factors of five speed up.

2. *Low-rank approximation of the Hessian* $\mathbf{H}$: To generate the samples we need fast evaluations of the inverse and inverse square-root of the Hessian. The former is related to the covariance and the latter is needed to "filter the noise". For lack of a better alternative, we follow Martin et al. (2012) and work with a low-rank approximation of the Hessian $\mathbf{H}$.

3. *Drawing samples*: With the above approximations for the Hessian, we generate N_{sample} samples $\{\mathbf{m}_i\}_{1\leq i\leq N_{\text{sample}}}$ according to the following proposal probability density:

$$q(\mathbf{m}) \propto \exp\left[-(\mathbf{m}-\mathbf{m}_{\text{MAP}})^T\mathbf{H}(\mathbf{m}-\mathbf{m}_{\text{MAP}})\right]. \tag{3}$$

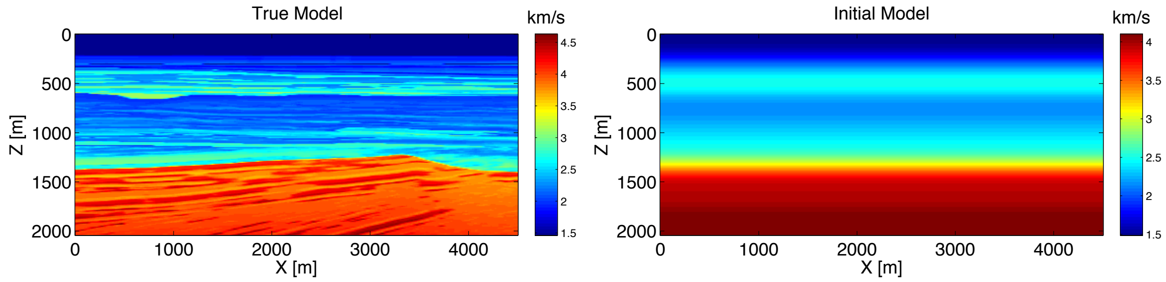
4. *Calculating* $\pi_{post}(\mathbf{m}_i)$: To decide to accept or reject samples, the posterior PDFs must be calculated for each sample (cf. Eq. 2). This is where we make an additional approximation by evaluating the data misfit over random subsets of sources. To avoid correlation in the error amongst different realizations for the velocity model, we use different subsets of sources for each realization. As we will see in the numerical examples below, this additional approximation does not decrease the quality of estimates for the posterior PDF.

5. *Metropolis-Hasting sample criterion:* For the actual decision to accept or reject a sample, we use Metropolist-Hasting criterion (Martin et al., 2012). This criterion involves the computation of $\alpha_i = \min\{1, \frac{\pi_{\text{post}}(\mathbf{m}_i)q(\mathbf{m}_{\text{accept}})}{\pi_{\text{pos t}}(\mathbf{m}_{\text{accept}})q(\mathbf{m}_i)}\}$ first, followed by drawing a random value for β_i from an uniform distribution $\mathscr{U}([0,1])$. Now, if $\alpha_i \geq \beta_i$ then we accept $\mathbf{m}_i$ and replace $\mathbf{m}_{\text{accept}}$ with $\mathbf{m}_i$, otherwise, we reject. By invoking this criterion, we make sure that we only accept samples that are consistent with the posterior distribution. The Markov Chain comes in because we base the decision on the most recently accepted sample of the velocity model only, i.e. $\mathbf{m} = \mathbf{m}_{\text{accept}}$. Following this procedure, we obtain an accepted subset of indices $\mathscr{I}_{\text{accept}}$ amongst the samples we have drawn.

6. *Estimating uncertainty parameters*: As a final step, we use the subset of indices that follow from the Metropolist-Hasting criterion to calculate the uncertainty parameters we are interested in, such as standard deviation, covariance and confidence intervals.

Numerical Examples

We applied the fast McMC method to estimate uncertainties associated with velocity models obtained via full-waveform inversion. After calculating the MAP estimate by carrying out conventional FWI, we computed a set of models using the procedure outlined above. The results of this exercise for the BG model, plotted in Figure 1(a), are included in Figure 3 and Figure 4. During the inversion, 15 equally-spaced frequencies ranging from 3 Hz to 17 Hz were included for 91 shots sampled at 50 meters and 451 receivers sampled at 10 meters. The maximum offset was 4.5 km. To start the inversion, we used the initial model plotted in Figure 1(b) that was obtained by smoothing the original model followed by a horizontal stack. As is custom, the inversion was carried out in consecutive (five) frequency bands using the previous result as a warm start. This gives us the MAP estimate plotted in Figure 3(a). For simplicity,



(a) True model (b) Initial model

Figure 1 *True model and initial model.*

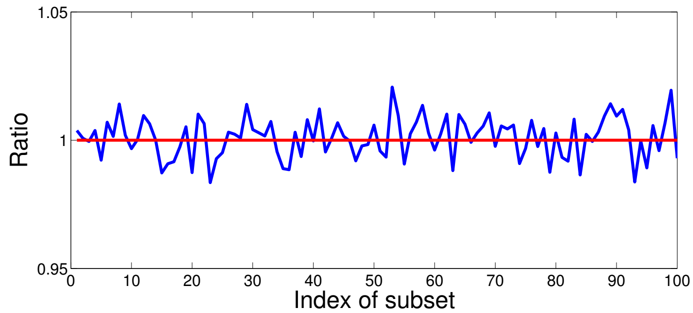


Figure 2 *Ratio of PDF using sub-set shots over PDF using all shots.*

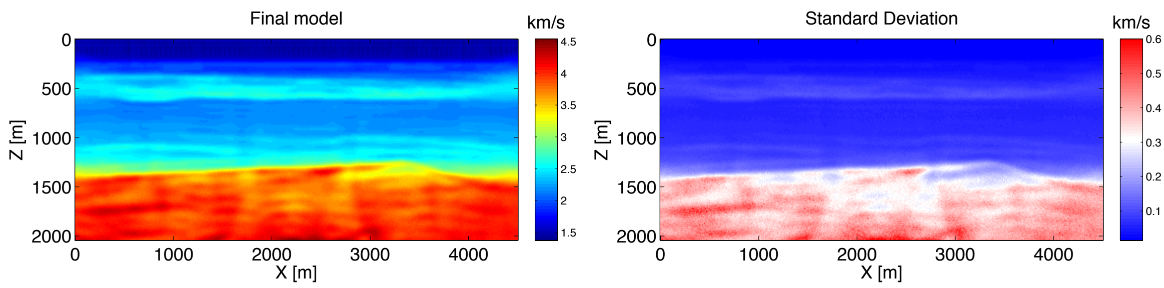
the uncertainty quantification was conducted for white Gaussian zero-centered noise with $\sigma = 0.05$ on the data and a white Gaussian covariance on the model that was centered around the initial model. The standard deviation on the model was taken to be 100 m/s.

To speed up the computation of the posterior distribution, we evaluated the data misfit for random subsets with 6 sources, which leads to a speedup of 15 X. To verify the accuracy of the approximation, we compared the posterior PDF at the MAP estimate for all data with the PDF computed with the random subsets only. We did this comparison for 100 different subsets and we computed the ratio between the true and approximate PDFs in Figure 2. It is clear from this figure that our approximation is accurate because the ratio is close to one for all random subsets.

To compute the uncertainty quantification, we generated 1000 random samples of which we accepted 384 samples with the Metropolis-Hasting criterion. The resulting estimates for the standard deviation and confidence intervals are shown for the MAP estimate (plotted in Figure 3(a)) in Figure 3(b) and Figure 4, respectively. The plots for the error bars are based on a 90 percent confidence interval and are plotted in Figure 4 for the three lateral positions (x = 540, 2040 and 4040 m). The true velocity and inversion result are represented with the red and black lines, respectively. The blue line represents the confidence interval for each point. From the standard deviation map (see Figure 3(b)), we observe that the velocity of the deep part has a larger standard deviation than that of the shallow part, as expected. A similar observation can be made for confidence interval plotted in Figure 4. Here, the probability density concentrates for shallow depths, while in deep areas, the probability density is less concentrated, which means that we are less confident.

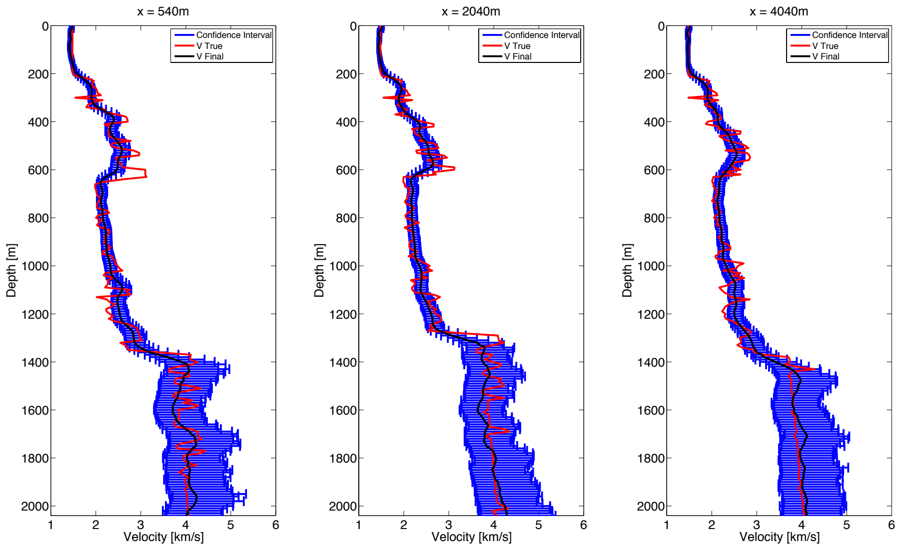
Conclusions

In this work, we propose a computationally tractable Markov chain Monte Carlo method where both the Hessian and the posterior distribution are approximated. The latter approximation is new and derives from recent findings on stochastic optimization, which reduce the number of wave-equation solves to



(a) Final model (b) Standard Deviation

Figure 3 *Final model and standard deviation.*



(a) x = 1 (b) x = 2 (c) x = 3

Figure 4 *Confidence intervals at x = 540, 2040, and 4040 m.*

evaluate the misfit term in the posterior distribution with a controllable error. While initial results are encouraging, several challenges remain such as the validity of the quadratic (Gaussian) approximation and the low-rank approximation of the Hessian.

Acknowledgements

We would like to thank all of the sponsors of our SINBAD project for their continued support and we also thank BG Group for providing the synthetic model we used in our example.

References

Bui-Thanh, T., Ghattas, O., Martin, J. and Stadler, G. [2013] A Computational Framework for Infinite-Dimensional Bayesian Inverse Problems Part I: The Linearized Case, with Application to Global Seismic Inversion. *SIAM Journal on Scientific Computing*, **35**(6), A2494–A2523.

Martin, J., Wilcox, L., Burstedde, C. and Ghattas, O. [2012] A Stochastic Newton MCMC Method for Large-scale Statistical Inverse Problems with Application to Seismic Inversion. *SIAM Journal on Scientific Computing*, **34**(3), A1460–A1487.

van Leeuwen, T. and Herrmann, F.J. [2013a] 3D Frequency-Domain Seismic Inversion with Controlled Sloppiness. Tech. rep.

van Leeuwen, T. and Herrmann, F.J. [2013b] Fast Waveform Inversion without Source-Encoding. *Geophysical Prospecting*, **61**, 10–19, ISSN 1365-2478.