Uncertainty quantification for Wavefield Reconstruction Inversion

Zhilong Fang[*], Chia Ying Lee[†], Curt Da Silva[*], Felix J. Herrmann[*], and Rachel Kuske[†]
[*]Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia
[†]Department of Mathematics, University of British Columbia

Abstract

In this work, we propose a method to quantify the uncertainty of wavefield reconstruction inversion under the framework of Bayesian inference. Unlike the conventional method using the wave equation as the forward mapping, we involve the wave equation misfit in the posterior distribution and propose a new posterior distribution. The negative log-likelihood of the new distribution is less nonlinear than that of the conventional posterior distribution, and its Gauss-Newton Hessian is a diagonal matrix that can be generated without any additional computational cost. We use the diagonal Gauss-Newton Hessian to derive an approximate Gaussian distribution at the maximum likelihood point to quantify the uncertainty. This method makes the uncertainty quantification for WRI computationally tractable and is able to provide reasonable uncertainty analysis based on our numerical results.

Introduction

Full-waveform inversion (FWI) (Tarantola and Valette, 1982a; Virieux and Operto, 2009) is a popular method in seismic exploration that uses the waveform information to reconstruct subsurface structure. However, conventional FWI suffers from nonlinearity and local minima due to the elimination of the PDE constraint in the optimization problem. Wavefield reconstruction inversion (WRI) (van Leeuwen and Herrmann, 2013; Peters et al., 2014) is a new approach of full-waveform inversion (FWI) that reduces the nonlinearity and lessens the effect of local minima. Instead of explicitly solving the wave equation, WRI relaxes the wave equation constraint by proposing a penalty formulation that explores both model and wavefield spaces. By exploiting this larger space, WRI reduces the likelihood of reaching local minima. Moreover, compared to conventional FWI, whose Gauss-Newton Hessian is typically a dense matrix, the corresponding Gauss-Newton Hessian of WRI is a diagonal matrix and can be generated without any additional cost once the wavefield is computed (van Leeuwen and Herrmann, 2013).

Statistical inversion aims to obtain the posterior distribution given the distribution of the observed data and prior model information (Tarantola and Valette (1982b)) and quantify the uncertainty based on the posterior distribution. However, quantifying the uncertainty for the conventional FWI is difficult. The forward mapping involving explicit PDE solves is very expensive computationally, which renders Markov chain Monte Carlo type methods intractable. Moreover, due to the strong nonlinearity of the forward mapping, it is also difficult to generate an approximate distribution for the conventional statistical FWI to correctly quantify the uncertainty.

In this work, we propose a new posterior distribution, in which we use the wave equation misfit of WRI. Compared to the conventional statistical FWI, the negative log-likelihood of this new posterior distribution is "less nonlinear" and the Gauss-Newton Hessian, which is diagonal, is easy to obtain. These properties allow us to be able to derive a Gaussian distribution that approximates the posterior distribution, and quantify uncertainty based on this approximate distribution. No additional computational cost related to sampling and estimating the Gauss-Newton Hessian is required, unlike for the standard FWI case. This makes the method computationally feasible to quantify the uncertainty of WRI. Numerical experiments illustrate that the uncertainty quantified by this method is reasonable.

Posterior Distribution for WRI

The deterministic wavefield reconstruction inversion is aimed at solving the following unconstrained optimization problem (1),

$$\min_{\mathbf{u},\mathbf{v}} \mathbf{F}(\mathbf{u},\mathbf{v}) = \sum_{k,l} \|\mathbf{P}_k\mathbf{u}_{k,l} - \mathbf{d}_{k,l}\|_2^2 + \lambda^2 \|\mathbf{A}_{k,l}(\mathbf{v})\mathbf{u}_{k,l} - \mathbf{q}_{k,l}\|_2^2, \tag{1}$$

which aims to balance the PDE misfit $\mathbf{A}(\mathbf{v})\mathbf{u}-\mathbf{q}$ with the data misfit $\mathbf{Pu}-\mathbf{d}$, simultaneously over velocity models $\mathbf{v}$ and wavefields $\mathbf{u}$, for all shots indexed by $k = 1,2,...,N_s$ and a few relevant frequencies indexed by $l = 1,2,...,N_f$. $\mathbf{P}$ is the operator that restricts wavefields to receiver locations, and $\mathbf{q}$ represents sources. This joint optimization problem, as stated, is difficult to solve and requires a large amount of memory to store all wavefields. One approach is to use the variable projection method (van Leeuwen and Herrmann, 2013) to reduce the number of parameters by finding optimal wavefields $\mathbf{u}$ for given velocity model $\mathbf{v}$ at each iteration. As a result, the original joint optimization problem (1) is reduced to the problem (2), and the optimal wavefield $\mathbf{u}$ is generated by solving the least-squares problem (3).

$$\min_{\mathbf{v}} \mathbf{F}(\mathbf{u}(\mathbf{v}),\mathbf{v}) = \sum_{k,l} \|\mathbf{P}_k\mathbf{u}_{k,l}(\mathbf{v}) - \mathbf{d}_{k,l}\|_2^2 + \lambda^2 \|\mathbf{A}_{k,l}(\mathbf{v})\mathbf{u}_{k,l}(\mathbf{v}) - \mathbf{q}_{k,l}\|_2^2, \tag{2}$$

$$\begin{pmatrix} \lambda\mathbf{A}_{k,l} \\ \mathbf{P}_k \end{pmatrix} \mathbf{u}_{k,l} = \begin{pmatrix} \lambda\mathbf{q}_{k,l} \\ \mathbf{d}_{k,l} \end{pmatrix} \tag{3}$$

In this work, we use the same idea from deterministic WRI to derive the posterior distribution based on the Bayesian inference. The conventional statistical FWI defines the posterior distribution as the product of data misfit likelihood and prior distribution, which only considers the uncertainty of data. However, the forward modeling kernel also should contain uncertainty due to the fact that we do not fit the PDE exactly. Based on this observation, we derive a new posterior distribution (4), which considers data misfit likelihood (the first term), wave equation misfit likelihood (the second term) and the prior distribution (the last term). Covariance matrices Σ_{noise}, Σ_{pde} and Σ_{prior} reflect uncertainty level and statistics of data, wave equation and prior model, respectively. Optimal wavefields $\mathbf{u}$ for fixed velocity model $\mathbf{v}$ are generated by solving the least-squares problem (5). By involving the wave equation misfit, the negative log-likelihood of eq. (4) is less nonlinear than that of conventional statistical FWI (van Leeuwen and Herrmann, 2013).

$$\rho_{\text{post}}(\mathbf{v},\mathbf{u}(\mathbf{v})) \propto \exp(-\sum_{k,l} \|\mathbf{P}\mathbf{u}_{k,l}(\mathbf{v}) - \mathbf{d}_{k,l}\|^2_{\Sigma^{-1}_{\text{noise}}} - \lambda^2 \|\mathbf{A}_{k,l}(\mathbf{v})\mathbf{u}_{k,l}(\mathbf{v}) - \mathbf{q}_{k,l}\|^2_{\Sigma^{-1}_{\text{pde}}} - \|\mathbf{v} - \mathbf{v}_{\text{prior}}\|^2_{\Sigma^{-1}_{\text{prior}}}) \quad (4)$$

$$\begin{pmatrix} \lambda \Sigma_{\text{pde}}^{-1/2}\mathbf{A}_{k,l} \\ \Sigma_{\text{noise}}^{-1/2}\mathbf{P}_k \end{pmatrix} \mathbf{u}_{k,l} = \begin{pmatrix} \lambda \Sigma_{\text{pde}}^{-1/2}\mathbf{q}_{k,l} \\ \Sigma_{\text{noise}}^{-1/2}\mathbf{d}_{k,l} \end{pmatrix} \quad (5)$$

Quantify the Uncertainty

The posterior probability distribution (4) is essential for the uncertainty quantification. However, quantifying statistical parameters based on (4) is extremely expensive because of the huge computational cost of solving least-squares problem (5) and the high dimension of $\mathbf{v}$. These properties renders traditional MCMC methods, which require an enormous amount of PDE solves (on the order of 10000 objective evaluations or more) in order to generate reasonable results, computationally intractable. An alternative method is to use a Gaussian distribution at the MAP point to approximate the original distribution, which is much cheaper to deal with than the posterior distribution (4).

The approximate Gaussian distribution for conventional FWI is difficult to generate because the Gauss-Newton Hessian of the negative log-likelihood of the posterior distribution is dense and requires a large number of PDE solves to calculate. On the other hand, the Gauss-Newton Hessian of the negative log-likelihood of (4) is a diagonal matrix. Both the gradient and the diagonal Gauss-Newton Hessian can be generated by eq. (6) and eq. (7) without additional computational cost, once we obtain wavefields $\mathbf{u}$. Using $\mathbf{g}$ and $\mathbf{H}$, we are able to derive a quadratic approximation of the negative log-likelihood. Fig. 1 shows the comparison of the negative log-likelihood (green line) and its quadratic approximation (blue line) at the MAP point in four different random directions. We can observe that around the MAP point, the quadratic approximation matches the negative logarithm function quite well, which implies that the approximate Gaussian distribution $\mathcal{N}(\mathbf{v}_{MAP} - \mathbf{H}^{-1}_{MAP}\mathbf{g}_{MAP}, \mathbf{H}^{-1}_{MAP})$ is a good approximation to the posterior distribution.

$$\mathbf{g} = \sum_{k,l} 2\lambda^2 \text{diag}(\text{conj}(\mathbf{u}_{k,l})) \frac{\partial \mathbf{A}_{k,l}}{\partial \mathbf{v}}^T \Sigma^{-1}_{\text{pde}} (\mathbf{A}_{k,l}\mathbf{u}_{k,l} - \mathbf{q}_{k,l}) \quad (6)$$

$$\mathbf{H} = \sum_{k,l} 2\lambda^2 \text{diag}(\text{conj}(\mathbf{u}_{k,l})) \frac{\partial \mathbf{A}_{k,l}}{\partial \mathbf{v}}^T \Sigma^{-1}_{\text{pde}} \frac{\partial \mathbf{A}_{k,l}}{\partial \mathbf{v}} \text{diag}(\mathbf{u}_{k,l}) \quad (7)$$

Numerical Experiment

We test our uncertainty quantification method for WRI on the BG compass model. The model size is 4.5km x 2km. The true model and initial model are shown in the Fig. 2. 91 sources are located at the surface every 50m and 451 receivers are located at the surface every 10m. Ten frequency bands ranging from {2,3,4}Hz to {29,30,31}Hz are used for the inversion. In this experiment, we assume that we know the noise free mean of the data, which may not be available in the practical case, and generate it by a 9-point finite difference method. The standard deviation for the noise of data is 1, the standard deviation for the PDE is 1 and the penalty parameter λ is selected to be 100. We assume that we do not have any prior information so that we do not include the prior distribution.

We first solve the deterministic optimization problem to find the MAP point (Fig. 3a). Then we calculate the gradient and Gauss-Newton Hessian at the MAP point and form the approximate Gaussian distribution.

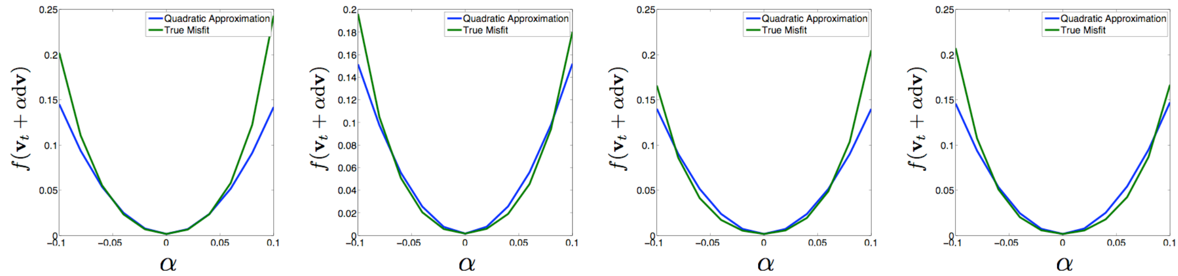
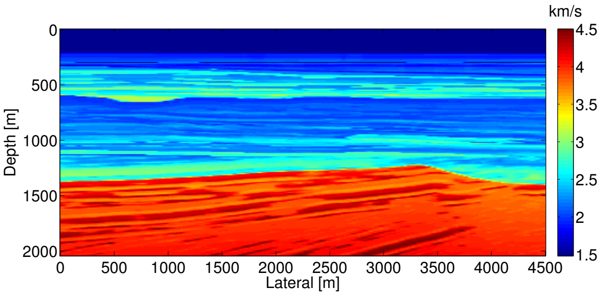
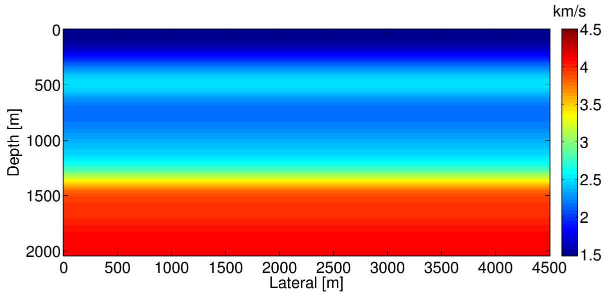


Figure 1: Comparisons of the true negative log-likelihood (green line) and the approximated quadratic function (blue line) in four different random directions d$\mathbf{v}$ with step-size α.



(a) True model

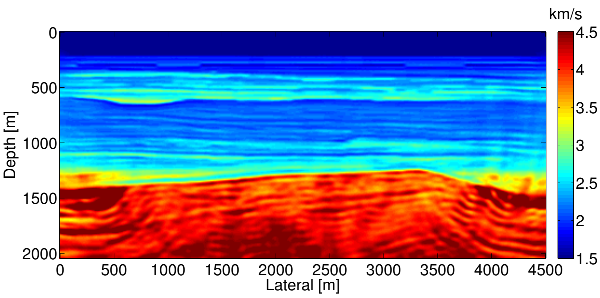


(b) Initial model

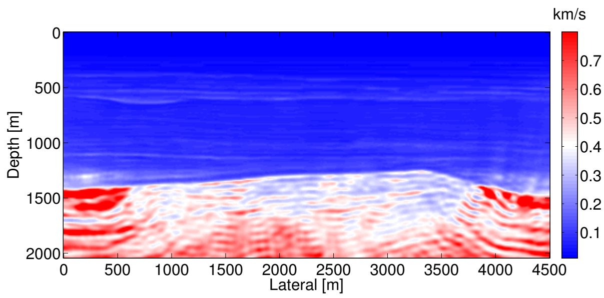
Figure 2: The true velocity model (a) and the initial velocity model (b).

The standard deviation (Fig. 3b) and confidence intervals (Fig. 4) are computed using our Gaussian approximation to the true distribution. From the computed standard deviation, we observe that the standard deviation at the deep and boundary areas of the model is higher than in other areas, which agrees with our intuition. We also observe similar phenomena in the confidence interval results that the velocity in deeper regions has a larger uncertainty compared to shallower regions. Both observations corresponds to the fact that the observed data is less sensitive to the velocity in these areas.

In order to verify our uncertainty results, we invert five sets of data generated by adding noise according to the prescribed noise distribution above. These five results are shown in the Fig. (4) with the initial model (blue line), true model (green line) and MAP result (black line). All five of these results lie in the confidence interval, which indicates that our results are reasonable.

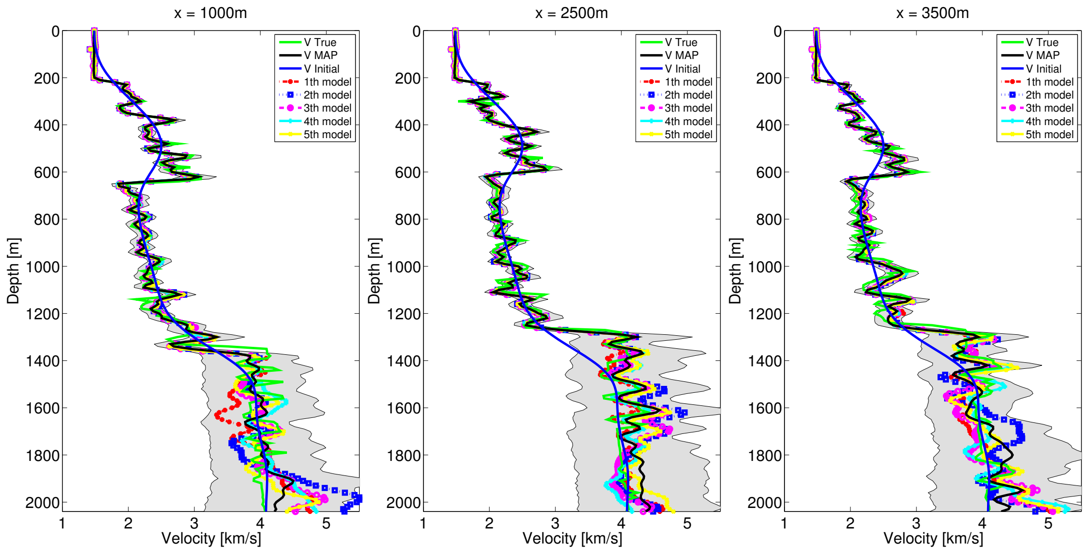


(a) Maximum likelihood point



(b) Standard Deviation

Figure 3: The Maximum likelihood point (a) and the standard deviation of the inversion result (b). At the deep part of the model and near the boundary, the velocity has large standard deviation since the observed data influences the inverted model much less in these regions than in the rest of the model. At the shallow part, the standard deviation is smaller, which shows less uncertainty at this area.



(a) x = 1000m (b) x = 2500m (c) x = 3500m

Figure 4: Confidence interval (gray background) at lateral position x = 1000m, 2500m, and 3500m. Green line - the true velocity, black line - the MAP result, blue line - the initial velocity and others - inversion result from noisy data obeying the specified noise distribution. All inversion results lie within the confidence interval we obtain.

Conclusion

In this work, we propose a method to quantify the uncertainty of wavefield reconstruction inversion by introducing the wave equation misfit to the posterior distribution. This makes the negative log-likelihood of the posterior distribution less nonlinear than in the FWI case and the diagonal Gauss-Newton Hessian is straightforward and efficient to generate. We use this Gauss-Newton Hessian to derive an approximate Gaussian distribution and quantify the uncertainty of the velocity model using this approximation. Since the Gauss-Newton Hessian is diagonal, we avoid solving a large number of PDEs that would otherwise have to be computed when estimating uncertainty for standard FWI. This method makes the uncertainty quantification for WRI computationally tractable and can provide reasonable uncertainty results based on the numerical experiments. Future work will focus on the situation that the noise free mean of data is not available.

Acknowledgment

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Statoil, Total SA, Sub Salt Solutions, WesternGeco, and Woodside.

References

Peters, B., Herrmann, F.J. and van Leeuwen, T. [2014] Wave-equation based inversion with the penalty method: adjoint-state versus wavefield-reconstruction inversion. *EAGE*, doi:10.3997/2214-4609.20140704.

Tarantola, A. and Valette, B. [1982a] Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics*, **20**(2), 219–232, ISSN 1944-9208, doi:10.1029/RG020i002p00219.

Tarantola, A. and Valette, B. [1982b] Inverse problems = quest for information. *Journal of Geophysics*, **50**, 159–170.

van Leeuwen, T. and Herrmann, F.J. [2013] Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, **195**, 661–667, doi:10.1093/gji/ggt258.

Virieux, J. and Operto, S. [2009] An overview of full-waveform inversion in exploration geophysics. *GEOPHYSICS*, **74**(6), WCC1–WCC26, doi:10.1190/1.3238367.