Uncertainty quantification for inverse problems with a weak wave-equation constraint

Zhilong Fang[1,*], Curt D. Silva[1], Rachel Kuske[2], Felix J. Herrmann[1]

[1]Department of EOAS, University of British Columbia, Vancouver, Canada
[2]Department of Mathmatics, University of British Columbia, Vancouver, Canada

*Email: fzlong88@gmail.com

Suggested members of the Scientific Committee

Abstract

In this work, we present a new posterior distribution to quantify uncertainties in solutions of wave-equation based inverse problems. By introducing an auxiliary variable for the wavefields, we weaken the strict wave-equation constraint used by conventional Bayesian approaches. With this weak constraint, the new posterior distribution is a bi-Gaussian distribution with respect to both model parameters and wavefields, which can be directly sampled by the Gibbs sampling method.

Keywords: Uncertainty, wave-equation

1 Introduction

In wave-equation based inverse problems, the goal is to infer the unknown model parameters from the observed data using the wave-equation as a constraint. Conventionally, the wave-equation is treated as a strict constraint in a Bayesian inverse problem. After eliminating this constraint, the problem involves the following forward modeling operator mapping model to predicted data:

$$f(\mathbf{m}) = \mathbf{P}\mathbf{A}(\mathbf{m})^{-1}\mathbf{q}, \qquad (1)$$

where the vectors $\mathbf{m} \in \mathbb{R}^{n_{\mathrm{grid}}}$ and $\mathbf{q} \in \mathbb{C}^{n_{\mathrm{grid}}}$ represent the discretized n_{grid}-dimensional unknown model parameters and known source term, respectively. The matrix $\mathbf{A} \in \mathbb{C}^{n_{\mathrm{grid}} \times n_{\mathrm{grid}}}$ represents the discretized wave-equation operator and the operator $\mathbf{P} \in \mathbb{R}^{n_{\mathrm{rcv}} \times n_{\mathrm{grid}}}$ projects the solution of the wave-equation $\mathbf{u} = \mathbf{A}(\mathbf{m})^{-1}\mathbf{q}$ onto the n_{rcv} receivers.

In the Bayesian framework, the solution of an inverse problem given observed data $\mathbf{d}$ is a posterior probability density function (PDF) $\rho(\mathbf{m}|\mathbf{d})$ expressed as [1]:

$$\rho(\mathbf{m}|\mathbf{d}) \propto \rho(\mathbf{d}|\mathbf{m})\rho(\mathbf{m}), \qquad (2)$$

where the likelihood PDF $\rho(\mathbf{d}|\mathbf{m})$ describes the probability of observing data $\mathbf{d}$ given model parameters $\mathbf{m}$ and the prior PDF $\rho(\mathbf{m})$ describes one's prior knowledge about the model parameters. Under the assumption that the noise in the data is Gaussian with zero mean and covariance matrix Γ_{n}, and the prior distribution is also Gaussian with a mean model $\tilde{\mathbf{m}}$ and covariance matrix Γ_{p}, the posterior PDF can be written as:

$$\rho(\mathbf{m}|\mathbf{d}) \propto \exp(-\frac{1}{2}(\|f(\mathbf{m})-\mathbf{d}\|_{\Gamma_{\mathrm{n}}^{-1}}^2 + \|\mathbf{m}-\tilde{\mathbf{m}}\|_{\Gamma_{\mathrm{p}}^{-1}}^2)). \qquad (3)$$

Due to the non-linear map $f(\mathbf{m})$ and the high dimensionality of the model parameters ($n_{\mathrm{grid}} \geq 10^5$), applying Markov chain Monte Carlo (McMC) methods to sample the posterior PDF (3) faces a difficult challenge of constructing a proposal PDF that provides a reasonable approximation of the target density with reasonable computational costs [2].

2 Posterior PDF with weak constraint

As the exact constraint $A(\mathbf{m})\mathbf{u} = \mathbf{q}$ leads to the difficulty of studying the corresponding posterior PDF, we weaken the constraint and arrive at a more generic posterior PDF with an auxiliary variable – wavefields $\mathbf{u}$ as follows:

$$\rho(\mathbf{m}, \mathbf{u}|\mathbf{d}) \propto \rho_1(\mathbf{d}|\mathbf{u})\rho_2(\mathbf{u}|\mathbf{m})\rho(\mathbf{m}), \text{ with}$$

$$\rho_1(\mathbf{d}|\mathbf{u}) \propto \exp(-\frac{1}{2}\|\mathbf{P}\mathbf{u} - \mathbf{d}\|_{\Gamma_{\mathrm{n}}^{-1}}^2), \text{ and}$$

$$\rho_2(\mathbf{u}|\mathbf{m}) \propto \exp(-\frac{\lambda^2}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{q}\|^2). \qquad (4)$$

Here the penalty parameter λ controls the trade off between the wave-equation and the data-fitting terms. As λ grows, the wavefields are more tightly constrained by the wave-equation. It is readily observed that the posterior PDF (2) is a special case of the posterior PDF (4) with $\rho_2(\mathbf{u}|\mathbf{m}) = 1$ and $\mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{q}$.

The new posterior PDF (4) has two important properties. First, the conditional PDF $\rho(\mathbf{u}|\mathbf{m}, \mathbf{d})$ of $\mathbf{u}$ on $\mathbf{m}$ is Gaussian. Second, if the matrix $\mathbf{A}(\mathbf{m})$ linearly depends on $\mathbf{m}$, the conditional PDF $\rho(\mathbf{m}|\mathbf{u}, \mathbf{d})$ of $\mathbf{m}$ on $\mathbf{u}$ is also Gaussian. Therefore, the posterior PDF (4) is bi-Gaussian with respect to $\mathbf{u}$ and $\mathbf{m}$.

3 Gibbs sampling

The bi-Gaussian property of the posterior PDF (4) provides a straight-forward intuition of applying Gibbs sampling method by alternatively drawing samples $\mathbf{u}$ and $\mathbf{m}$ from the conditional PDFs $\rho(\mathbf{u}|\mathbf{m}, \mathbf{d})$ and $\rho(\mathbf{m}|\mathbf{u}, \mathbf{d})$. At the kth iteration of the Gibbs sampling, starting from the point $\mathbf{m}_k$, the conditional PDF $\rho(\mathbf{u}|\mathbf{m}_k, \mathbf{d}) = \mathcal{N}(\overline{\mathbf{u}}, \mathbf{H}_u^{-1})$ with the Hessian matrix $\mathbf{H}_u$ and the mean wave-fields $\overline{\mathbf{u}}$ given by:

$$\mathbf{H}_u = \mathbf{P}^\top \Gamma_{\mathrm{p}}^{-1} \mathbf{P} + \lambda^2 \mathbf{A}(\mathbf{m}_k)^\top \mathbf{A}(\mathbf{m}_k),$$
$$\overline{\mathbf{u}} = \mathbf{H}_u^{-1}(\mathbf{A}(\mathbf{m}_k)^\top \mathbf{q} + \mathbf{P}^\top \Gamma_{\mathrm{p}}^{-1}\mathbf{d}). \quad (5)$$

To draw a sample $\mathbf{u}_{k+1} \sim \mathcal{N}(\overline{\mathbf{u}}, \mathbf{H}_u^{-1})$, we first compute the Cholesky factorization of $\mathbf{H}_u = \mathbf{L}_u^\top \mathbf{L}_u$, where $\mathbf{L}_u$ is an upper triangular matrix. Then we apply $\mathbf{L}_u$ to compute $\overline{\mathbf{u}}$ and $\mathbf{u}_{k+1}$ by:

$$\overline{\mathbf{u}} = \mathbf{L}_u^{-1}\mathbf{L}_u^{-\top}(\mathbf{A}(\mathbf{m}_k)^\top \mathbf{q} + \mathbf{P}^\top \Gamma_{\mathrm{p}}^{-1}\mathbf{d}),$$
$$\mathbf{u}_{k+1} = \overline{\mathbf{u}} + \mathbf{L}_u^{-1}\mathbf{r}_u, \mathbf{r}_u \sim \mathcal{N}(0, \mathbf{I}). \quad (6)$$

With $\mathbf{u}_{k+1}$, the conditional PDF $\rho(\mathbf{m}|\mathbf{u}_{k+1}, \mathbf{d}) = \mathcal{N}(\overline{\mathbf{m}}, \mathbf{H}_m^{-1})$. The mean $\overline{\mathbf{m}} = \mathbf{m}_k - \mathbf{H}_m^{-1}\mathbf{g}_m$ with gradient and Hessian expressed as:

$$\mathbf{g}_m = \mathbf{G}^\top(\mathbf{A}(\mathbf{m}_k)\mathbf{u}_{k+1} - \mathbf{q}) + \Gamma_{\mathrm{p}}^{-1}(\mathbf{m}_k - \tilde{\mathbf{m}}),$$
$$\mathbf{H}_m = \mathbf{G}^\top \mathbf{G} + \Gamma_{\mathrm{p}}^{-1}, \quad (7)$$

where $\mathbf{G} = \frac{\partial \mathbf{A}(\mathbf{m})\mathbf{u}_{k+1}}{\partial \mathbf{m}}$ is the sparse Jacobian matrix. The new sample $\mathbf{m}_{k+1}$ can be computed by:

$$\mathbf{m}_{k+1} = \overline{\mathbf{m}} + \mathbf{L}_m^{-1}\mathbf{r}_m, \mathbf{r}_m \sim \mathcal{N}(0, \mathbf{I}), \quad (8)$$

where $\mathbf{L}_m = \mathbf{H}_m^{1/2}$.

4 Numerical experiment

We apply our proposal method to an inverse problem constrained by the Helmholtz equation. We use single source, single frequency data to invert a 1D gridded squared slowness profile $m(z) = 1/(v_0 + \alpha z)^2$ for values $v_0 = 2000\,\mathrm{m/s}$ and $\alpha = 0.75$. We use frequency increments of $1\,\mathrm{Hz}$, a grid spacing of $50\,\mathrm{m}$, a maximum offset of $10000\,\mathrm{m}$, and a maximum depth of $5000\,\mathrm{m}$. Both source and receivers are located at the surface of the model. Gaussian noise with covariance matrix $\Gamma_{\mathrm{n}} = \mathbf{I}$ is added to the data. The mean model of the prior distribution is selected to have $v_0 = 2000\,\mathrm{m/s}$ and $\alpha = 0.65$. The covariance matrix is set to $\Gamma_{\mathrm{p}} = 2 \times 10^{-8}\mathbf{I}$. We use the Gibbs sampling method to generate 10^6 samples with $\lambda = 10^5$. We compare the true

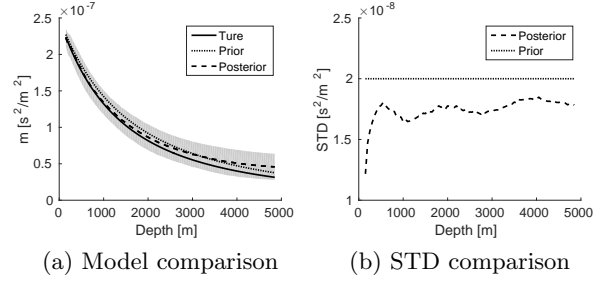

(a) Model comparison   (b) STD comparison

Figure 1: (a) Comparison of true model (solid), mean model of prior PDF (dotted) and posterior PDF (dashed); (b) Comparison of STDs of prior (dotted) and posterior (dashed) PDFs.

model, the prior mean model, and the posterior mean model in Figure 1a, and the prior and posterior standard deviations (STD)in Figure 1b. The 90% confidence interval of the posterior distribution is also shown in Figure 1a (gray background). Compared to the prior distribution, the posterior mean model has a smaller error in the shallow part but a larger error in the deep part. Meanwhile, the posterior STD in the shallow part has a larger decrease from the prior STD than that in the deep part. Both facts implies that data has a larger influence on the squared slowness in the shallow part than that in the deep part.

5 Conclusion

By weakening the wave-equation constraint in a controlled manner, we arrive at a novel formulation of posterior PDF for wave-equation inversion. This new posterior PDF is bi-Gaussian with respect to model parameters and wave-fields. Numerical experiment demonstrates that this posterior PDF can be successfully and efficiently sampled by the Gibbs sampling method.

6 Acknowledgements

References

[1] A. Tarantola, and B. Valette, Inverse Problems = Quest for Information, *Journal of Geophysics* **50** (1982), pp. 159-170.

[2] J. Martin, L. C. Wilcox, C. Burstedde, and O. Ghattas, A stochastic newton MCMC method for large-scale statistical inverse problems with application to seismic inversion *SIAM Journal on Scientific Computing* **34** (2012), pp. A1460-A1487