Bayesian Physics-Informed Neural Networks for seismic tomography based on the eikonal equation

Rongxi Gou*, Yijie Zhang, Xi’an Jiaotong University and Xueyu Zhu, The University of Iowa

SUMMARY

The high cost of acquiring sufficient amount of seismic data for training has limited the use of machine learning in seismic tomography. To overcome this issue, in this work, eikonal equation is used as a prior knowledge in the learning framework to establish the constraint relationship between inverse velocity and reconstruct traveltimes. Besides, to mitigate the effects due to noisy data and quantify the uncertainty in the prediction, Bayesian Physics-Informed Neural Networks (BPINN) is used and a general Bayesian inference algorithm Stein Variational Gradient Descent (SVGD) is used to facilitate the training process. The proposed method was tested on a synthetic velocity model with different noise levels in the data. The results show the velocity model can be estimated accurately and the traveltimes can be well approximated. The predictive uncertainty offered by BPINN can provide reasonable uncertainty assessments in the predictive results.

INTRODUCTION

Seismic tomography is one of the most effective methods to study the structure of the underground interface, among which traveltome maps are a widely used method for subsurface model building. Seismic tomography can be typically solved by minimizing the misfit of the traveltimes and simulated traveltimes generated by synthetic velocity models. As an inverse problem, seismic tomography is always suffered by the serious nonlinearity and strong dependence on the initial model.

Recently, machine learning have become extremely popular thanks to its adaptive learning and approximation power. However, the performance of purely data-driven machine learning approach is largely dependent on the training data which could lead to poor performance in sparse/noisy datasets. Many recent efforts in the scientific machine learning community have focused on these challenges. Notably, Physics-Informed Neural Networks (PINN) (Raissi et al., 2019) have been shown to address a series of problems caused by small dataset, in which the underlying governing equations are introduced as a regularization term into the loss function. For seismic application, based on eikonal equation, PINN has been used to construct the traveltome field from sparse observed traveltimes (Smith et al., 2020; Waheed et al., 2021b) and solve the corresponding inverse problems to predict the velocity field (Waheed et al., 2021a). In addition, PINN is leveraged to handle the full waveform inversion based on acoustic wave equation (Rashti-Behesht et al., 2021).

However, the uncertainty in the prediction due to the scariness and noises in the data is less discussed in the conventional machine learning methods and the vanilla PINN. To overcome the issue, Bayesian approaches has been introduced to provide the uncertainty of the output results. Notably, the so called Bayesian Physics-Informed Neural Networks (BPINN) (Sun and Wang, 2020; Yang et al., 2021; Ceccarelli, 2021; Smith et al., 2022) uses physics equation to provide additional prior knowledge while consider the parameters and output of the neural networks in a probabilistic way to provide predictive uncertainty.

In this work, we investigate the predictive capability of the BPINN in the seismic tomography based on the eikonal equation. A synthetic velocity model is used to demonstrate the performance of the proposed method under different noise levels.

THEORY

Eikonal equation

Seismic tomography is an important approach to characterize the structure of the underground, it supplies a reasonable initial model for the seismic inversion and migration. The eikonal equation is widely used in seismic tomography (Waheed et al., 2021a). To overcome the instability of the singular points, the factorized eikonal equation is used:

\[
\mathcal{L}(x, x) = \begin{cases} \nabla \tau(x)^2 - \frac{1}{\nu^2(x)} = 0, & \forall x \in \Omega, \\ \tau(x) - 1 = 0. & \end{cases}
\]

where \( \tau(x) \) represents the traveltimes from the source point \( x_s \) to any point in the physical domain \( \Omega \), \( \nu(x) \) is the velocity defined in \( \Omega \). In this work, our goal is to inverse velocity model and reconstruct the traveltimes from sparse sampled velocity and traveltimes.

Bayesian Neural Network

Bayesian Neural Network (Blundell et al., 2015), which considers the network parameters as random variables, can not only gives the predicted results but also the uncertainty quantification of the output results. According to Bayesian theorem, the posterior distribution of the network output can be got through equation (2):

\[
P(\theta|\mathcal{D}) = \frac{P(\mathcal{D}|\theta)P(\theta)}{P(\mathcal{D})} \sim P(\mathcal{D}|\theta)P(\theta),
\]

where \( \theta, \mathcal{D} \) in (2) represent the network parameters and datasets, respectively. Evaluating the integral in (2) can be intractable for high dimensional problems. To address this issue, Markov chain Monte Carlo (MCMC) (Andrieu et al., 2003), Variational Inference (VI) (Graves, 2011; Blei et al., 2017) have been adopted to approximate the posterior distribution. However, MCMC can be infeasible for extreme
high dimensional problems and the performance of VI is limited by the predefined family of distributions to match target posterior distribution ([Blei et al., 2017]). In this work, Stein Variational Gradient Descent (SVGD) ([Liu and Wang, 2016]), a natural approach of gradient descent for full Bayesian inference is used to approximate $P(\theta|\mathcal{D})$ by minimizing the KL divergence as follows:

$$KL(q(\omega)||p(\omega)) = \int q(\omega) \log \frac{q(\omega)}{p(\omega)} d\omega,$$

(3)

where $q(\omega)$ and $p(\omega)$ are densities over the set of parameters $\omega$. $q(\omega)$’s distribution is defined by parameters $\theta$ to approximate target distribution $p(\omega)$. SVGD employs a set of particles $\{\theta_i\}_{i=1}^n$ to approximate the target probability distribution in order to minimize the KL divergence. During an iteration, we suppose that $\theta = \theta + \epsilon \phi(\theta)$, where $\epsilon$ means the step size, $\phi(\theta)$ represents the direction of parameters upgrading. In Liu’s work ([Liu and Wang, 2016]), (3) is proved to descend fastest when $\epsilon$ is satisfied:

$$\nabla_{\theta} KL(q||p) = -E_{\theta \sim q} D_{KL}(q||p),$$

(4)

$$\nabla_{\theta} \phi(\theta) = \nabla_{\theta} \log p(\theta) \phi(\theta)^T + \nabla_{\theta} \phi(\theta).$$

With the gradient direction, we can use the SVGD algorithm to upgrade particles as [5]:

$$\phi(\theta) = \frac{1}{n} \sum_{j=1}^{n} [k(\theta_j, \theta)\nabla_{\theta_j} (\log p(\theta_j) + \nabla_{\theta_j} k(\theta_j, \theta))],$$

$$\theta_{j}^{t+1} = \theta_j^t + \epsilon \phi(\theta_j^t),$$

(5)

where $t$ represents the number of iterations, $k(\cdot, \cdot)$ represents a positive kernel function. In our work, radial basis function (RBF) (6) is used:

$$k(\theta_j, \theta) = \exp\left(-\frac{||\theta - \theta_j||^2}{2l^2}\right),$$

(6)

where $l$ is the median distance between particles $\{\theta_i\}_{i=1}^n$ to control the bandwidth of RBF.

Bayesian physics-informed neural network

For the seismic tomography problem based on (1), given a set of $n_d$ measurements of velocity and traveltime factor:

$$\{(\tau(x_i^t, x^t), v(x_i^t))\}_{i=1}^{n_d},$$

We use two independent multi-layer neural networks $\mathcal{N}_v$ and $\mathcal{N}_\tau$ to approximate velocity and reconstructed traveltime factor respectively:

$$\hat{\tau}(x, x) = \mathcal{N}_\tau(x, x; \theta_\tau),$$

$$\hat{v}(x) = \mathcal{N}_v(x; \theta_v)(v_{\text{max}} - v_{\text{min}}) + v_{\text{min}}.$$

(7)

To accelerate networks converge, we normalized velocity data through

$$v_n(x) = \frac{v(x) - v_{\text{min}}}{v_{\text{max}} - v_{\text{min}},}$$

where $v_{\text{max}}$ and $v_{\text{min}}$ are the maximum and minimum velocity value on domain area $\Omega$.

To take in account the underlying physics (1), the log likelihood can be expressed as (9):

$$\log(\text{likelihood}) = \log p(\mathcal{D}|\theta, \Sigma_{\mathcal{D}}) + \log p(\mathcal{D}|\theta, \Sigma_{\mathcal{D}}),$$

$$\log p(\mathcal{D}|\theta, \Sigma_{\mathcal{D}}) \propto \left(-\frac{1}{2\Sigma_{\mathcal{D}}} \sum_{t=1}^{n_d} (\hat{v}(x_t^t) - v(x_t^t))^2 + \frac{n_d}{2} \log(\frac{1}{2\Sigma_{\mathcal{D}}})\right),$$

$$\log p(\mathcal{D}|\theta, \Sigma_{\mathcal{D}}) \propto \left(-\frac{1}{2\Sigma_{\mathcal{D}}} \sum_{t=1}^{n_d} (\hat{\tau}(x_t^t, x^t) - \tau(x_t^t, x^t))^2 + \frac{n_d}{2} \log(\frac{1}{2\Sigma_{\mathcal{D}}})\right),$$

(9)

where $p(\mathcal{D}|\theta, \Sigma_{\mathcal{D}})$ measures the error between predicted and input data in a probabilistic way. $p(\mathcal{D}|\theta, \Sigma_{\mathcal{D}})$ evaluates the likelihood of physics constraint $\mathcal{D}$, which is the residual of (1). $n$ is the number of residual points. $\Sigma_{\mathcal{D}}$ is the covariance matrix of Gaussian distribution reflecting our confidence in physics constraint. In our work, data are contaminated with Gaussian white noise. To approximate the data uncertainty, the prediction error of $\mathcal{N}_v$ and $\mathcal{N}_\tau$ are assumed to follow Gaussian distribution. Trainable parameters $\Sigma_{\theta_v}$ and $\Sigma_{\theta_\tau}$ estimate the covariance matrix of the predefined distribution $p(\mathcal{D}|\theta, \Sigma_{\mathcal{D}})$. The whole algorithm of this work is shown in Algorithm 1.

### Algorithm 1: Bayesian physics-informed neural network based on Stein Variational Gradient Descent

**Input:** A set of sparse, noisy data of traveltime and velocity.

**Output:** Suitable neural networks parameters to reconstruct traveltime and inverse velocity in domain area.

**Initialization:** Sample prior distributions for $\theta$ and $\Sigma_{\mathcal{D}}$ with $n$ particles;

for $t=0$ to $t_{\text{end}}$ do

1. Calculate $L(\theta) = \log p(\theta) + \log p(\Sigma_{\mathcal{D}}) + \log(\text{likelihood})$;

2. Calculate $\phi(\theta) = \frac{1}{n} \sum_{j=1}^{n} [k(\theta_j, \theta)\nabla_{\theta} L(\theta) + \nabla_{\theta} k(\theta_j, \theta)];$

3. Upgrade $\theta_{j}^{t+1} = \theta_j^t + \epsilon \phi(\theta_j^t)$ and networks parameters $\theta$;

end

### EXPERIMENTS

In this section, we demonstrate the capability of the proposed method on a synthetic velocity model. A $1 \times 5$ km$^2$ synthetic velocity model shown in Figure 1 is used for experiment. The grid size of the model is 0.02 km spacing in the x, y-axis, so a $51 \times 251$ grid is used to characterize the velocity model. The red and black dots in Figure 1 represent the locations of the receivers and sources respectively. 101 receivers picked up traveltime from 11 sources. $11 \times 101$ groups of traveltime data and velocity data are treated as input for BPINN. These sample
BPINN for tomography

points are used as supervised data in the training. 5% random Gaussian noise is added to the measured data of the equispaced receivers and sources.

All the grid points are used to evaluate the misfit of the eikonal equation and the uncertainty of $p(\theta | R, \Sigma)$. The CPU toolkit scikit-fmm [the scikit-fmm team 2021] is used to calculate the travelt ime from a source point to receivers located in model space.

The two independent multilayer neural network $N_v$ and $N_\tau$ both contain 10 hidden layers. $N_v$ has 10 neurons per layer with SELU activation function and the output of $N_v$ has been normalized with a sigmoid function. $N_\tau$ has 20 neurons per layer with Swish activation function [Sun and Wang 2020; Waheed et al. 2021a]. The trainable parameters $\Sigma_D$ and $\Sigma_D$ to fit uncertainty of the neural networks output are initialized by a series of identically distributed inverse Gamma distribution $IG(\theta | \alpha, \beta)$ with $\alpha = 2$ and $\beta = 1 \times 10^{-6}$. Besides, probabilistic particles $\theta$ is assumed to follow a student’s t distribution $St(\theta | n)$ with $n = 2$. The Adam optimizer with a batch size of 1024 for 1000 epochs is used with learning rate of $1 \times 10^{-3}$ to upgrade particles $\theta$ and learning rate of $1 \times 10^{-4}$ to upgrade $\Sigma_\theta$. The experiment was implemented by modifying an open source BPINN library [Sun and Wang 2020] on NVIDIA GeForce RTX 2080 Ti GPU and the training time is about 10 hours.

To estimate the predicted velocity, absolute relative error (ARE) is used,

$$\text{ARE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{v}(x_i) - v(x_i)|}{v(x_i)}$$

(10)

$\hat{v}(x_i)$ and $v(x_i)$ are the predicted and ground truth velocity respectively.

The inverted velocity model by BPINN is shown in Figure 2. The ARE of the output is 0.0136, indicating the networks has recovered the velocity profile perfectly under noisy sparse data. Notably, the arch structure of the subsurface velocity profile between 2.0 to 3.0 km is well approximated.

Figure 3 shows the corresponding absolute relative error and uncertainty represents by one standard deviation given by BPINN. There is no infallible reference to estimate quantitative uncertainty, but we can judge the rationality of the given uncertainty. The uncertainty is large at the bottom left and right corners which means low confidence. This is consistent with the the larger discrepancy of the true and predicted velocity model, where are far from labeled data. As a result, larger uncertainty presents. The exact values of the uncertainty may vary if alternative methods are used, but the distribution is generally consistent, demonstrating predictive capability of BPINN.

The performance of the BPINN based on data with different noise levels is further investigated in Table 1. The uncertainty represents by the one standard deviation has been normalized by the corresponding true velocity data, and the mean and maximum values of the uncertainty about predicted results are list in this table. As expected, the ARE and uncertainty increases as the noise level increases.

CONCLUSION

In this work, the BPINN is introduced for seismic tomography based on eikonal equation. SVGD algorithm is used to enable the efficient training. The predictive capability of the proposed method is demonstrated via a synthetic velocity model are used for testing the feasibility of the proposed method. To investigate the capability of BPINN on uncertainty quantification, the input datasets are imported with different levels of Gaussian noise. The predicted velocity model and the corresponding quantitative uncertainty given by BPINN demonstrate that the proposed scheme offers a new way for seismic tomography.

ACKNOWLEDGMENTS

R. Gou and Y. Zhang would like to thank National Natural Science Foundation of China under Grant 42174137 and the China Postdoctoral Science Foundation funded project under Grant 2021M692521. X. Zhu was supported by the Simons Foundation (504054).
BPINN for tomography

Figure 1: The ground truth velocity model. There are 11 equispaced sources on $z = 0$ km. 51 receivers uniformly space on $z = 0$ km with an interval distance of 0.1 km. Besides, 50 equispaced receivers on $x = 2.5$ km with a uniform spacing of 20 m.

Figure 2: The predicted velocity model of the BPINN with a set of noisy data with 5% noise.

Figure 3: (a) The misfit (measured by ARE) between true velocity model and predicted model. (b) The standard deviation of the predicted velocity model of BPINN.
Table 1: Error and uncertainty of the predicted velocity model based on BPINN with different levels of Gaussian noise. Here, mean/max std represent the mean and maximum values of the uncertainty of predicted results, respectively.

| Level of noise | ARE  | Velocity uncertainty (mean/max std) |
|----------------|------|------------------------------------|
| 5%             | 0.0136 | 0.0292/0.0440                     |
| 15%            | 0.0146 | 0.0725/0.1071                     |
| 25%            | 0.0151 | 0.1003/0.1471                     |
BPINN for tomography

REFERENCES

Andrieu, C., N. De Freitas, A. Doucet, and M. J. Jordan, 2003, An introduction to MCMC for machine learning: Machine learning, 50, 5–43.

Baker, N., F. Alexander, T. Bremer, A. Hagberg, Y. Kevrekidis, H. Najm, M. Parashar, A. Patra, J. Sethian, S. Wild, et al., 2019, Workshop report on basic research needs for scientific machine learning: Core technologies for artificial intelligence: Technical report, USDOE Office of Science (SC), Washington, DC (United States).

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe, 2017, Variational inference: A review for statisticians: Journal of the American Statistical Association, 112, 859–877.

Blundell, C., J. Cornebise, K. Kavukcuoglu, and D. Wierstra, 2015, Weight uncertainty in neural network: International conference on machine learning, PMLR, 1613–1622.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe, 2017, Variational inference: A review for statisticians: Journal of the American Statistical Association, 112, 859–877.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe, 2017, Variational inference: A review for statisticians: Journal of the American Statistical Association, 112, 859–877.

Brooks, S., A. Gelman, G. Jones, and X.-L. Meng, 2011, Handbook of markov chain monte carlo: CRC press.

Ceccarelli, D., 2021, Bayesian physics-informed neural networks for inverse uncertainty quantification problems in cardiac electrophysiology.

Graves, A., 2011, Practical variational inference for neural networks: Advances in neural information processing systems, 24.

Izzathullah, M., U. B. Waheed, I. R. Yildirim, and T. Alkhalifah, 2022, Predictive uncertainty quantification for bayesian physics-informed neural network (PINN) in hypocentre estimation problem: 83rd EAGE Conference and Exhibition 2022, European Association of Geoscientists & Engineers, 1–5.

Liu, Q., and D. Wang, 2016, Stein variational gradient descent: A general purpose bayesian inference algorithm: Advances in neural information processing systems, 29.

Lu, L., X. Meng, Z. Mao, and G. E. Karniadakis, 2021, DeepXDE: A deep learning library for solving differential equations: SIAM Review, 63, 208–228.

Psaros, A. F., X. Meng, Z. Zou, L. Guo, and G. E. Karniadakis, 2022, Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons: arXiv preprint arXiv:2201.07766.

Raisi, M., P. Perdikaris, and G. E. Karniadakis, 2019, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations: Journal of Computational Physics, 378, 686–707.

Rasht-Behesht, M., C. Huber, K. Shukla, and G. E. Karniadakis, 2021, Physics-informed neural networks (PINNs) for wave propagation and full waveform inversions: arXiv preprint arXiv:2108.12035.

Smith, J. D., K. Azizzadenesheli, and Z. E. Ross, 2020, Eikonet: Solving the eikonal equation with deep neural networks: IEEE Transactions on Geoscience and Remote Sensing, 59, 10685–10696.

Smith, J. D., Z. E. Ross, K. Azizzadenesheli, and J. B. Muir, 2022, HypoSVI: Hypocentre inversion with stein variational inference and physics informed neural networks: Geophysical Journal International, 228, 698–710.

Sun, L., and J.-X. Wang, 2020, Physics-constrained bayesian neural network for fluid flow reconstruction with sparse and noisy data: Theoretical and Applied Mechanics Letters, 10, 161–169.

Rasht-Behesht, M., C. Huber, K. Shukla, and G. E. Karniadakis, 2021, Physics-informed neural networks (PINNs) for wave propagation and full waveform inversions: arXiv preprint arXiv:2108.12035.

Smith, J. D., K. Azizzadenesheli, and Z. E. Ross, 2020, Eikonet: Solving the eikonal equation with deep neural networks: IEEE Transactions on Geoscience and Remote Sensing, 59, 10685–10696.

Smith, J. D., Z. E. Ross, K. Azizzadenesheli, and J. B. Muir, 2022, HypoSVI: Hypocentre inversion with stein variational inference and physics informed neural networks: Geophysical Journal International, 228, 698–710.

Sun, L., and J.-X. Wang, 2020, Physics-constrained bayesian neural network for fluid flow reconstruction with sparse and noisy data: Theoretical and Applied Mechanics Letters, 10, 161–169.

the scikit-fmm team, 2021, scikit-fmm, a CPU toolkit to formulate the travel time. (https://pythonhosted.org/scikit-fmm/)

Waheed, U. b., T. Alkhalifah, E. Haghighat, C. Song, and J. Virieux, 2021a, PINNtomo: Seismic tomography using physics-informed neural networks: arXiv preprint arXiv:2104.01588.

Waheed, U. b., E. Haghighat, T. Alkhalifah, C. Song, and Q. Hao, 2021b, PINNeik: Eikonal solution using physics-informed neural networks: Computers & Geosciences, 155, 104833.

Yang, L., X. Meng, and G. E. Karniadakis, 2021, B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data: Journal of Computational Physics, 425, 109913.