A Stable Neural Network-Based Eikonal Tomography using Hard-Constrained Measurements

Mohammad Taufik[1], Tariq Alkhalifah[1], and Umair Waheed[2]

[1]King Abdullah University of Science and Technology (KAUST), Physical Sciences and Engineering Division, Thuwal, Saudi Arabia
[2]King Fahd University of Petroleum and Minerals

March 26, 2023

Abstract

Eikonal tomography, or travel time inversion, has been one of the primary seismological tools for decades and has been used to understand Earth's properties and dynamic processes. At the heart of the inversion process is the need for an accurate, and preferably flexible, eikonal solver to compute the travel time field. Most of the conventional eikonal solvers, however, suffer from first-order convergence errors and difficulties in dealing with irregular computational grids. Physics-informed neural networks (PINNs) have been introduced to tackle these problems and have shown great success in addressing those challenges. Nevertheless, these approaches still suffer from slow convergence and unstable training dynamics due to the multi-term nature of the loss function. To improve on this, we propose a new formulation for the isotropic eikonal equation, which imposes boundary conditions as hard constraints. We employ the theory of functional connections to the eikonal tomography problem, which allows for the utilization of a single loss term for training the PINN model. Through rigorous numerical tests, its efficiency, stability, and flexibility in tackling a variety of cases, including topography-dependent and 3D models, are attested, thus providing an efficient and stable PINN-based eikonal tomography.

A Stable Neural Network-Based Eikonal Tomography using Hard-Constrained Measurements

Mohammad H. Taufik [1*], Tariq Alkhalifah[1], and Umair bin Waheed [2]

[1]Physical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, 23955, Saudi Arabia
[2]Department of Geosciences, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia

Key Points:

- We develop a novel implementation of eikonal tomography via physics-informed neural networks (PINNs), in which we embed the measurements (data) as hard constraints into the eikonal equation.
- We demonstrate the efficiency and flexibility of the new formulation for different use cases (no modifications to the eikonal), which marks a significant improvement from the conventional finite-difference based methods.
- The new formulation yields a single, stable, objective function removing the burden of figuring out an appropriate weighting for the loss terms of the previous PINN-based approaches. The algorithm also demonstrates its stability in handling complex 3D velocity distributions.

*Physical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, 23955, Saudi Arabia

Corresponding author: Mohammad Taufik, mohammad.taufik@kaust.edu.sa

Abstract

Eikonal tomography, or travel time inversion, has been one of the primary seismological tools for decades and has been used to understand Earth's properties and dynamic processes. At the heart of the inversion process is the need for an accurate, and preferably flexible, eikonal solver to compute the travel time field. Most of the conventional eikonal solvers, however, suffer from first-order convergence errors and difficulties in dealing with irregular computational grids. Physics-informed neural networks (PINNs) have been introduced to tackle these problems and have shown great success in addressing those challenges. Nevertheless, these approaches still suffer from slow convergence and unstable training dynamics due to the multi-term nature of the loss function. To improve on this, we propose a new formulation for the isotropic eikonal equation, which imposes boundary conditions as hard constraints. We employ the theory of functional connections to the eikonal tomography problem, which allows for the utilization of a single loss term for training the PINN model. Through rigorous numerical tests, its efficiency, stability, and flexibility in tackling a variety of cases, including topography-dependent and 3D models, are attested, thus providing an efficient and stable PINN-based eikonal tomography.

Plain Language Summary

Machine learning techniques have proven to be beneficial in various scientific and engineering applications. One of these techniques, termed as physics-informed neural networks (PINNs), attempts to solve parametric partial differential equations (PDEs) by imposing the necessary conditions and PDE residuals into the objective function being optimized. The multi-component nature of the PINNs' objective function, however, often renders the training to be unstable. Here, we attempt to circumvent this pathology by reformulating the PDE (eikonal equation) such that the necessary conditions (travel time measurements) are naturally included in the PDE. Thus, the objective function yields a single term that stabilizes the training process. We report that the proposed formulation is also flexible to be incorporated into different eikonal tomography applications from 2D to 3D Earth's model seamlessly.

1 Introduction

Seismic tomography (inversion) has been and still is one of the primary seismological tools for understanding the properties and description of the dynamic processes inside the Earth. Pioneered by the work of Dziewonski et al. (1977) and Aki et al. (1977), eikonal tomography facilitates the tomographic process by inverting the measured travel time information of the seismic waves. Although limited by the underlying ray theoretic high-frequency asymptotic assumption, decades of development and utilization justify the usability of the method in a wide range of seismological applications. In global seismology, eikonal tomography is regularly used to locate earthquakes (Thurber, 1983; Klein, 2002; Hauksson et al., 2012), image the Earth's interior via body waves (H. Zhang & Thurber, 2003), surfaces waves (Lin et al., 2007, 2009; Lin & Ritzwoller, 2011), joint surface-body waves tomography (Rawlinson & Fishwick, 2012; Obrebski et al., 2012; Fang et al., 2016), and deriving global velocity models (Simmons et al., 2012, 2021). On the exploration front, it is heavily involved in the imaging of shallow crust (J. Zhang & Toksöz, 1998; Zelt & Barton, 1998; Zelt et al., 2006), velocity model building for migration (Marsden, 1993; Dessa et al., 2004; Bergman et al., 2004) and full-waveform inversion (Virieux & Operto, 2009; Fichtner, 2010; Tavakoli et al., 2017), and reservoir monitoring and characterization (Bording et al., 1987).

At the heart of the travel time inversion process lies the eikonal equation. The eikonal equation is a first-order nonlinear partial differential equation (PDE) that can be derived, for example, from the infinite-frequency assumption of the wave equation. The tomographic algorithms can be performed by either performing linearization to the inversion

operator (e.g., via ray theory (Červený, 2000) or via the eikonal (Alkhalifah, 2002)) or using the adjoint-state method (Leung & Qian, 2006; Plessix, 2006; Taillandier et al., 2009). The conventional linearization approach gives rise to the need of computing (and storing) the Fréchet matrix, and its inverse, which is a challenge for 3D dense measurements. The adjoint-state approach avoids the calculation of such a matrix and has shown promising results over the years. Shared across these algorithms is the need for an efficient and preferably flexible way to solve the eikonal equation. To achieve this, conventional algorithms resort to either the ray-based approaches (Červený, 2000; Julian et al., 1977; Um & Thurber, 1987) or the grid-based approaches (Vidale, 1988; Qin et al., 1992; Sethian, 1996; Rawlinson & Sambridge, 2004; Zhao, 2005). However, most of these algorithms still suffer from the so-called first-order convergence error and sharp velocity changes; in these scenarios, the algorithms may fail to converge (Rawlinson et al., 2008). More importantly, a significant modification (e.g., coordinate transformation) needs to be incorporated when dealing with irregular surface topography.

On another front, employing physics-informed neural networks (PINNs) (Raissi et al., 2019) as a replacement to the conventional eikonal solver has shown promising results in addressing these issues. Smith et al. (2021) and Waheed et al. (2021b) demonstrated the more accurate travel time fields produced by the PINN-based solver for travel time modeling and identify the mesh-independent nature of such solvers. Taufik et al. (2022) demonstrated the nonlinear interpolation ability of the PINN-based solver in dealing with non-uniformly sampled data. Moreover, Waheed et al. (2021a) proposed a framework for treating the ill-posed body wave tomography problems by simultaneously utilizing two neural networks (NNs) to invert for the travel time and velocity fields. Chen et al. (2022) extended the framework to invert for a surface wave tomography application.

These successful cases can partially be explained by a well-known nonlinear interpolation ability of the NNs (Hornik et al., 1989) as well as the use of travel time factorization approaches. Travel time factorization is introduced to overcome the so-called source-singularity problem (Fomel et al., 2009). The problem statement then becomes, given an initial (background) travel time field, the NN is trained to predict a variable that maps the initial to the actual travel time field. Therefore, the factorization implicitly drives the PINNs' loss function to include at least two terms, namely the PDE residual and the boundary-related condition loss. For travel time tomography, both the boundary condition and data mismatch need to be explicitly imposed during the training process of PINNs. Hence, on top of the already challenging training dynamics, the multi-term nature of the PINNs' training under this regime often induces more instability due to the need to properly balance the loss terms with respect to their weight and the number of samples.

Several previous works attribute the PINNs' instability to improper loss function sampling. Wang et al. (2022) reformulated the PINNs' objective functions by explicitly obeying the physical causality. McClenny and Braga-Neto (2020) proposed the use of an efficient weighting scheme to the multi-term objective function. Based on this idea, (Yu et al., 2022) utilized an effective sampling based on the residual values. In this work, orthogonal to the aforementioned approaches, we develop a new formulation for the isotropic eikonal equation by imposing the boundary conditions as hard constraints (HC). We implement the theory of functional connections (TFC) (Schiassi et al., 2020) into the eikonal tomography problem, which admits a single loss term for training the PINNs-based model. Trained in this fashion, not only does our formulation still inherit all the known favorable properties of PINNs, but it also makes the training more robust and efficient. These properties are the result of the natural inclusion of boundary conditions in the optimization problem. Combining this with a data interpolation NN makes the proposed scheme handle sparse and irregularly sampled measurements accurately.

Thus, the contributions of our work can be summarized as follows,

1. Introduce a novel PINN-based eikonal tomography framework that inverts for the velocity and travel time fields with a single loss term for training the PINNs model.
2. Derive a new isotropic eikonal equation with hard-constrained data measurements.
3. Propose the use of a nonlinear data interpolation scheme to handle sparse and irregularly sampled measurements.
4. Analyze different travel time factorization approaches that best suit PINN-based eikonal tomography.
5. Provide, to the best of our knowledge, the first evidence of the use of PINN-based eikonal tomography to handle regional-scale 3D data.

In the following, we begin by providing the necessary theoretical background to derive the proposed framework, followed by several numerical experiments to showcase the flexibility and efficiency of the framework in different acquisition scenarios.

2 Methodology

In this section, we will first present the governing eikonal equation in its factorized form. We will then discuss how to embed the measured travel time data as hard constraints in the eikonal equation. Finally, we share the rationale for using a neural network to perform nonlinear data interpolation.

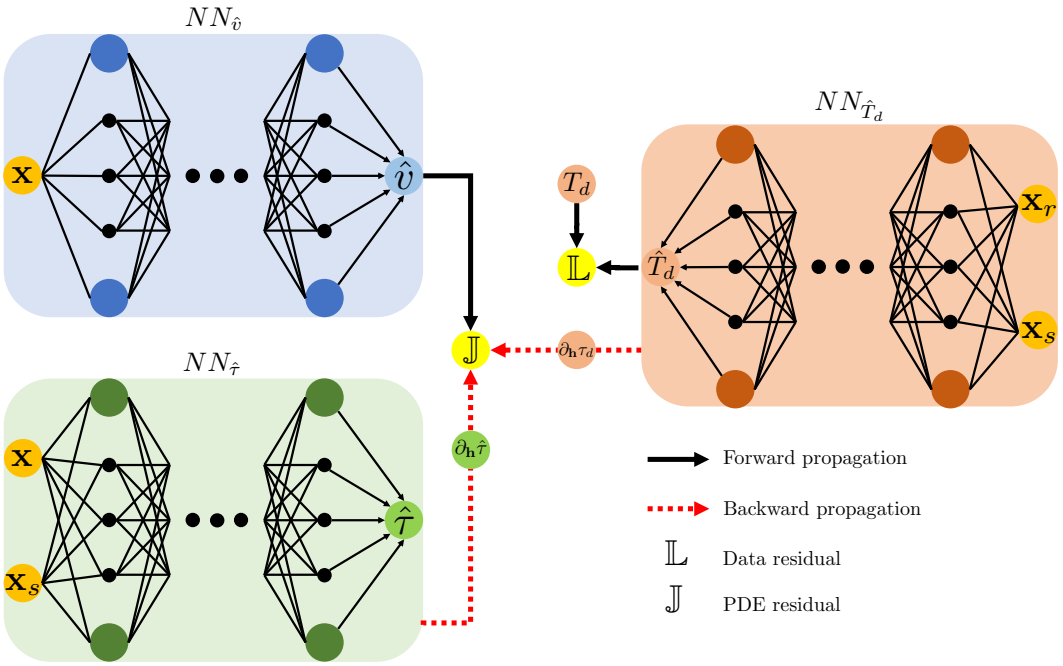

Figure 1. The proposed workflow for our PINN-based eikonal tomography problem.

2.1 The Isotropic Eikonal and Its Factorization

The eikonal equation for an isotropic medium can be written in the form of (Červený, 2000):

$$|\nabla T(\mathbf{x})|^2 = \frac{1}{v^2(\mathbf{x})}, \tag{1}$$

where T denotes the travel time field and v denotes the medium phase velocity, both as a function of position vector $\mathbf{x} = \{x, y, z\} \in \mathbb{R}^3$ for a three-dimensional Cartesian sys-

tem. To mitigate the source singularity problem, the travel time field can also be split into an additive form, which yields the so-called factored eikonal equation given by:

$$T(\mathbf{x}) = \tau(\mathbf{x}) + T_0(\mathbf{x}), \tag{2}$$

where a scalar τ is introduced to map the background travel time T_0 to the actual travel time T. Specifically, the background travel time is simply defined as:

$$T_0(\mathbf{x}) = \frac{\sqrt{(\mathbf{x} - \mathbf{x}_s)^2}}{v_0(\mathbf{x})}, \tag{3}$$

which is given by the distance from the source location $\{\mathbf{x}_s\}$ divided over some background velocity, e.g., a constant velocity (v_0).

Using equation 2 to represent the travel time field T, we have,

$$\nabla T(\mathbf{x}) = \nabla\tau(\mathbf{x}) + \nabla T_0(\mathbf{x}). \tag{4}$$

2.2 Incorporating Data as Hard Constraint

To incorporate the measured travel time data in the eikonal equation as a hard constraint, we use the theory of connection functions (Schiassi et al., 2020). For this derivation, we first consider a typical surface tomography experiment in which the data measurements T_d take place along a constant depth surface at z_r. In this case, we suggest the following representation of the travel time field:

$$T(\mathbf{x}) = \zeta(z)\hat{\tau}\left(\mathbf{x}\right) + \tau_d\left(x, y\right) + T_0\left(\mathbf{x}\right), \tag{5}$$

where

$$\zeta(z) = z - z_r, \tag{6}$$

and

$$\tau_d(x, y) = T_d(x, y) - T_0(x, y, z = z_r). \tag{7}$$

The term $\hat{\tau}(\mathbf{x})$ is parameterized by a neural network (NN) functional. Alternatively, for a typical cross-hole measurement, where the source and receiver locations are fixed at a certain lateral location, we can easily adapt equation 5 using

$$\zeta(x) = x - x_r. \tag{8}$$

More generally, we can even further impose a topography-dependent recording surface in which the z_r in equation 6 is a function of x (for a surface tomography problem). Compared to the original factored eikonal equation, here the new factor $\hat{\tau}$ is not guaranteed to be positive everywhere in the domain of interest.

To this end, the overall workflow of our proposed PINN-based tomography is demonstrated in Figure 1. The core of the inversion is shown by the left two NNs while the right NN corresponds to the optional data fitting/interpolation NN ($NN_{\hat{T}_d}$) trained prior to the main inversion (training) process. The PINNs tomography model consist of two NNs, trained simultaneously from randomly initialized weights, to predict the velocity ($NN_{\hat{v}}$) and the travel time ($NN_{\hat{\tau}}$). The inputs to the travel time NN function are the location of the source, $\{\mathbf{x}_s\}$ and the position in space $\{\mathbf{x}\}$, whereas only $\{\mathbf{x}\}$ are inputs to the velocity NN function. Using automatic differentiation (AD) (Baydin et al., 2018), we can compute the lateral travel time derivatives ($\partial_{\mathbf{h}}\hat{\tau}$). The data NN function of the source and sensor locations is trained in a supervised fashion prior to the PINNs' training using a loss function, $\mathbb{L}$, that measures the misfit between the predicted data and the measured data T_d. The data network takes as input the location of the source $\{\mathbf{x}_s\}$ and receiver $\{\mathbf{x}_r\}$. Once trained, the data travel time NN will not only provide us with the ability to handle sparse measurements, it also admits the gradient of the travel time data

through AD ($\partial_{\mathbf{h}}\hat{\tau}_d$), which is needed for the new eikonal as we will see. We can compute the gradient either prior or during the PINNs training.

Using equation 4, the travel time gradient components of the eikonal can be formulated as:

$$\frac{\partial T\left(\mathbf{x}\right)}{\partial x} = \zeta(z)\frac{\partial \hat{\tau}\left(\mathbf{x}\right)}{\partial x} + \frac{\partial \tau_d\left(x,y\right)}{\partial x} + \frac{\partial T_0\left(\mathbf{x}\right)}{\partial x}, \tag{9}$$

$$\frac{\partial T\left(\mathbf{x}\right)}{\partial y} = \zeta(z)\frac{\partial \hat{\tau}\left(\mathbf{x}\right)}{\partial y} + \frac{\partial \tau_d\left(x,y\right)}{\partial y} + \frac{\partial T_0\left(\mathbf{x}\right)}{\partial y}, \tag{10}$$

and

$$\frac{\partial T\left(\mathbf{x}\right)}{\partial z} = \zeta(z)\frac{\partial \hat{\tau}\left(\mathbf{x}\right)}{\partial z} + \frac{\partial \zeta(z)}{\partial z}\hat{\tau}\left(\mathbf{x}\right) + \frac{\partial T_0\left(\mathbf{x}\right)}{\partial z}. \tag{11}$$

Hence, the proposed loss function for the PINN, $\mathbb{J}$, can be constructed by plugging in the gradients from the travel time network (equations 9 to 11) and the velocity $\hat{v}$ from the velocity network into equation 1.

$$\mathbb{J} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial T\left(\mathbf{x}_i\right)}{\partial x}^2 + \frac{\partial T\left(\mathbf{x}_i\right)}{\partial y}^2 + \frac{\partial T\left(\mathbf{x}_i\right)}{\partial z}^2 - \frac{1}{\hat{v}^2(\mathbf{x}_i)}\right). \tag{12}$$

2.3 Non-linear Data Interpolation

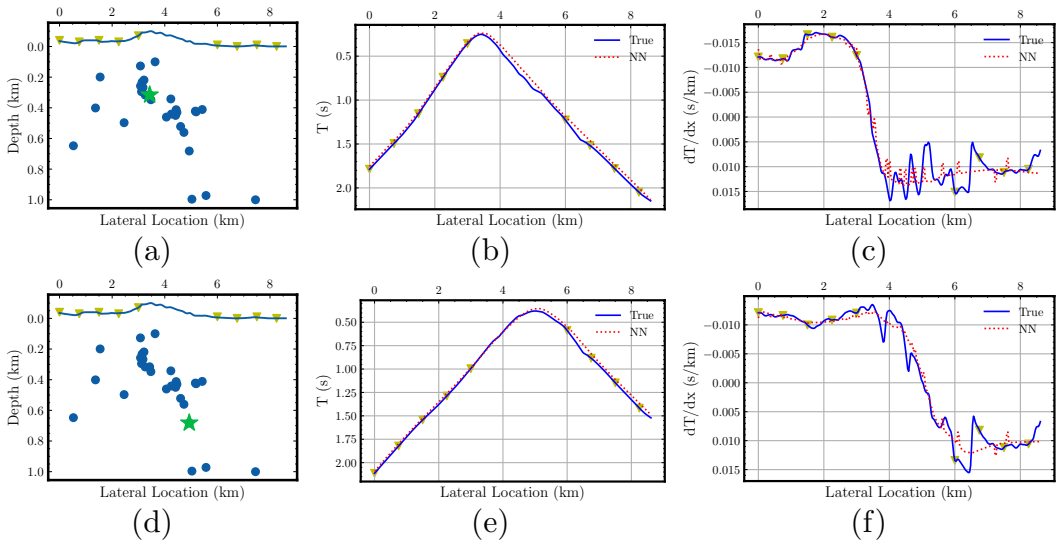

Figure 2. The data NN interpolation demonstration. The acquisition setup (left column, a and d). Travel time data and its horizontal derivatives (middle and right columns). The events used in the training are depicted by the blue dots, while results in the middle and right columns correspond to the green star event, and the receivers are denoted by the yellow triangles. True corresponds to those travel times and horizontal derivatives evaluated by solving the eikonal equation numerically on a fine grid.

One of the most celebrated and well-known fact about NN is that it acts as a good universal function approximator (Hornik et al., 1989). Thus, it naturally provides us with multidimensional (determined by the number of input coordinates) nonlinear interpolation. Since the data NN learns the travel time for multiple sources and receivers characterized by their 3D location in space, the deep nature of the network function allows it to utilize most of the travel time information to predict travel times for sources and

receivers not represented in the training (interpolation). This feature is crucial to its effectiveness in the tomography application.

To exemplify our point, consider two measured data from two different event locations (Figures 2) in which the data are recorded along a certain mountain top. The data are synthetically generated using the Marmousi model (Figure 5a). The receiver arrays are distributed almost regularly, but with a large gap near the mountain top (around 3.5 km location). The sources (given by the blue dots) are sampled from the earthquake events provided by the catalog in (Fang et al., 2022), and are the ones used to train the data NN. The middle column in Figure 2 shows the predicted travel times, compared to the eikonal solution, at all points on the surface for a source given by the green star, top and bottom rows separately. We can see that the NN managed to capture the apex of the hyperbolic travel time curve quite accurately–even when no receivers are present in this area. More importantly, the lateral derivatives (Figure 2, c and f) predicted (by performing backpropagation on the NN) are also in good agreement with the one evaluated directly from the numerical solution of the eikonal solution using fine sampling.

3 Numerical Experiments

In this section, we examine the flexibility and stability of the proposed single-term PINNs' loss function (equation 12). The first subsection will cover the evaluation of the method in a 2D exploration seismological setting where the seismometers are placed along a borehole with varying data sparsity. To further demonstrate the flexibility of the formulation, we dedicate the second subsection for different recording surface settings in a surface tomography setup on a synthetic 2D velocity model. We will also demonstrate, in the third subsection, numerical reasoning behind the use of additive as opposed to multiplicative factorization for the travel time. Finally, we showcase the stability of the method in handling complex 3D velocity distribution. A single graphics processing unit NVIDIA Quadro RTX 8000 (48 GB memory) is used to perform the inversion (Data NN and PINNs' training).

3.1 Crosswell Tomography Experiment in 2D Medium

In this setting, we use a cropped 2D 1x1 km^2 SEG Advanced Modeling Program (SEAM) velocity model depicted in Figure 3a. We use a 0.01 km grid spacing in both (lateral and vertical) directions. The data are synthetically generated using a fast-marching-based eikonal solver and recorded at $x = 1$ km. We utilize an NN with 10 hidden layers having 10 neurons in each layer for the velocity NN and the data NN and a 20 hidden layers with 10 neurons in each layer for the travel time NN. All the three networks are initialized using the Glorot uniform initialization (Glorot & Bengio, 2010). The data NN is trained prior to the PINNs training using 500 epochs. The PINNs model is trained for 1000 epochs using the generated 112,211 training samples ($\mathbf{x} = \{x, z\}$) with a single (full) batch optimization. All the NNs use the Exponential Linear Unit (ELU) activation function for the hidden layers, the rectified linear unit (ReLU) is used on the last neuron of the velocity NN. The training of all data NN and the PINN is done with an Adam optimizer (Kingma & Ba, 2014). We use an initial learning rate of 0.00015 and decrease its value by half every 100 training epochs.

To highlight the ability of this framework in handling sparse data measurements, we share three scenarios differing in the receiver spacing. We consider three different receiver spacing of 0.02, 0.06, and 0.15 km. The sources are located (along the vertical z-direction) at x=0 km with a vertical spacing of 0.1 km starting at z=0 km. Figures 3 and 4 offer the cross-sectional view and vertical profiles for the different scenarios. From the inverted velocity section (Figure 3), we observe that PINNs are able to invert for the velocity accurately. Shown in Figure 3, we observe, as expected, a slight degradation with increasing receiver spacing in the inverted velocity's lateral variation. This degradation,

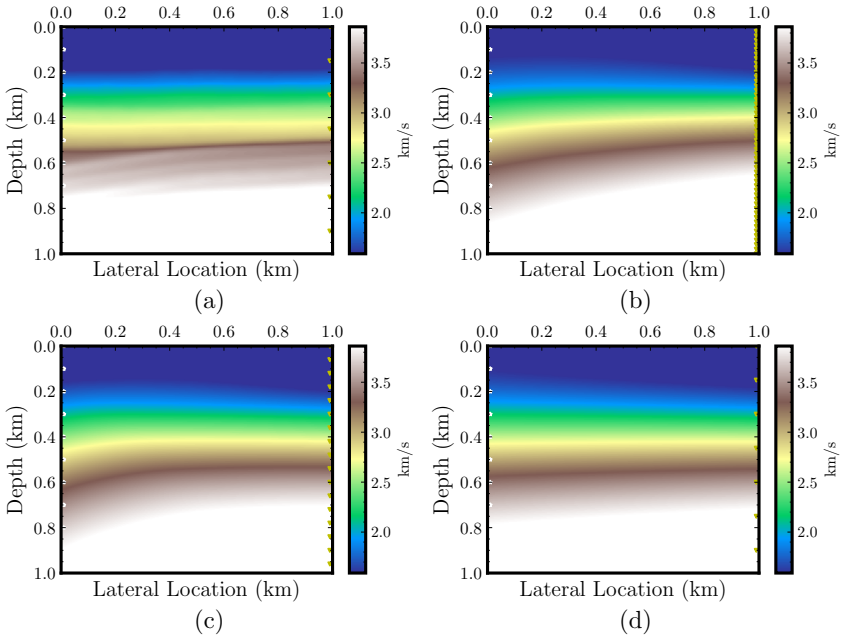

Figure 3. Inverted velocity sections for different receivers spacing (0.02, 0.06, and 0.15 km) in (b-d) comnpared to the the true velocity (a). The sources (white stars) and receivers (yellow triangles) are located at a fixed lateral location of 0 km and 1 km, respectively.

however, seems negligible as the inverted velocity on the largest receiver spacing still captures the shallow low-velocity layers of the model (Figure 4c) and follows the general increasing with depth velocity trend.

3.2 Surface Seismic Tomography in a 2D Medium

To further demonstrate the flexibility of the new travel time formulation (equation 5), we consider four distinct surface tomography tests with different data acquisition setups. These include the typical regularly sampled dense shot-receiver geometry, the same sampling scheme as in the first scenario with a gap, sparse shot distribution, and topography-dependent layout in a passive seismic recording setup. In all of these tests, we use an NN with 10 hidden layers containing 10 neurons in each layer for the velocity network and the data interpolation network. An NN with 20 hidden layers having 20 neurons in each layer is used for the higher dimensional travel time NN. We train the data NN for 5000 epochs. The travel time and velocity NNs use the ELU activation function while the data NN uses a rectified linear unit (ReLU) activation function. All of the three NNs use an Adam optimizer. We utilize a portion of the Marmousi model (Figure 5a) with a maximum lateral location of 8.6 km and maximum depth of 1 km. We discretize the computational domain using a 10 m vertical and 30 m lateral grid spacing.

Having the shot-receiver pairs regularly and densely sampled at a constant-depth acquisition setup, we consider the tomogram from the first scenario (Figure 5b) to be the reference solution for the other three scenarios. Specifically, the sources and receivers are sampled at a spacing interval of 200 m and 20 m, respectively. As shown in Figure 5b, the inversion manages to capture the general lateral variation accurately. We then use a sparser receiver sampling of 300 m and introduce a gap (in the sources and receivers, like an obstacle) that extends from 1.2 to 5.8 km lateral location in the model. Depicted in Figure 5c is the corresponding inverted velocity. We see degradation in terms of the

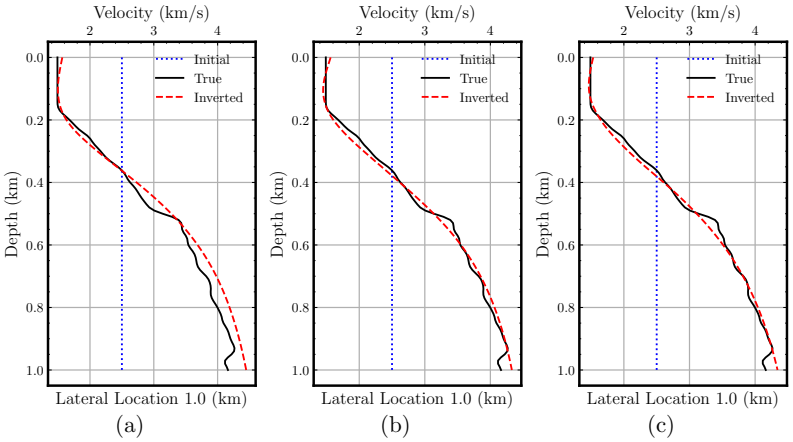

Figure 4. Vertical velocity profiles comparison for the cross-hole experiment for different receivers spacing (0.02, 0.06, and 0.15 km). Profiles in (a-c) are extracted at the borehole location (1 km lateral location) from Figure 3.

general lateral velocity resolution, especially under the gap, as we reduce almost half of the recording surface. The degraded tomogram, however, might still be considered a reasonable initial velocity model for further imaging (e.g., full waveform inversion). Using receivers sampled densely like in the first scenario, we reduce the shot sampling interval to 2.7 km in the third scenario. From the inverted velocity profile (Figure 5d), we can see that even with only four shots, the inverted velocity model captures the lateral variation of the actual model with high fidelity. Although some deep structures are obviously not well resolved, compared to the denser shot distribution (Figure 5b), the two tomograms are in good agreement.

To further highlight the efficacy of the new PINNs formulation, we test the same problem with a topography-dependent surface recording. It is straightforward to derive the PINNs' objective function by introducing a different ζ in equation 6, i.e., $\zeta(x, z)$. We also demonstrate the framework's ability to handle uneven source distributions by considering sampled earthquake locations from the southern part of California (Fang et al., 2022). The overall improvement from the tomogram can be attributed to the fact that now we image the transmission from the source as opposed to the diving waves in the previous three scenarios.

Finally, vertical velocity profiles are shown to further analyze the reconstructed velocity models. Although in general, the vertical profiles are identical (at least at the given lateral location), we can see that the 7-source experiments (Figures 5d and 6c) produce identical results when compared to the dense measurements (Figures 5b and 6a). The loss curves for the different scenarios are shown in Figure 7. We obtain slightly faster convergence with less data.

3.3 Suitable Factorization Approach for Data Hard-Constrained Eikonal Tomography

Before proceeding to the last sets of experiments, it is worth noting the role of travel time factorization on the PINNs' inversion. Specifically, the choice of factorization plays a major role in ensuring stability of the inversion. To this end, most of the previous works suggest the use of multiplicative factorization (i.e., changing the addition operation into

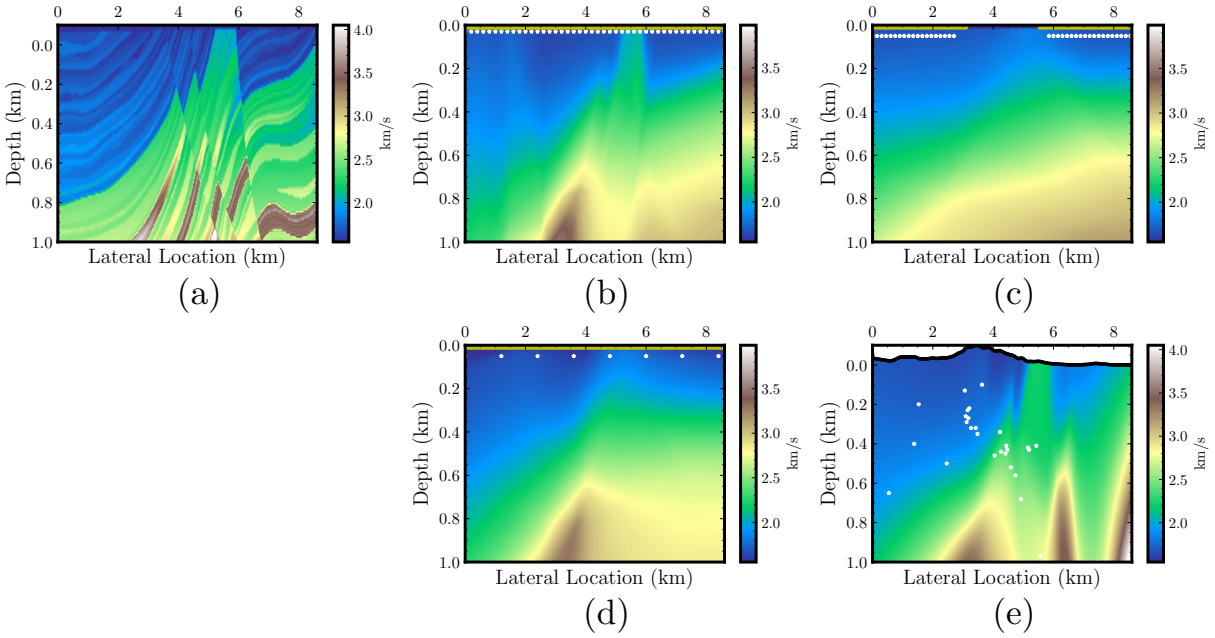

Figure 5. Inverted velocity sections for four different data acquisition scenarios (b-e) compared to the true velocity (a). The sources and receivers are denoted by the white and yellow dots, respectively.

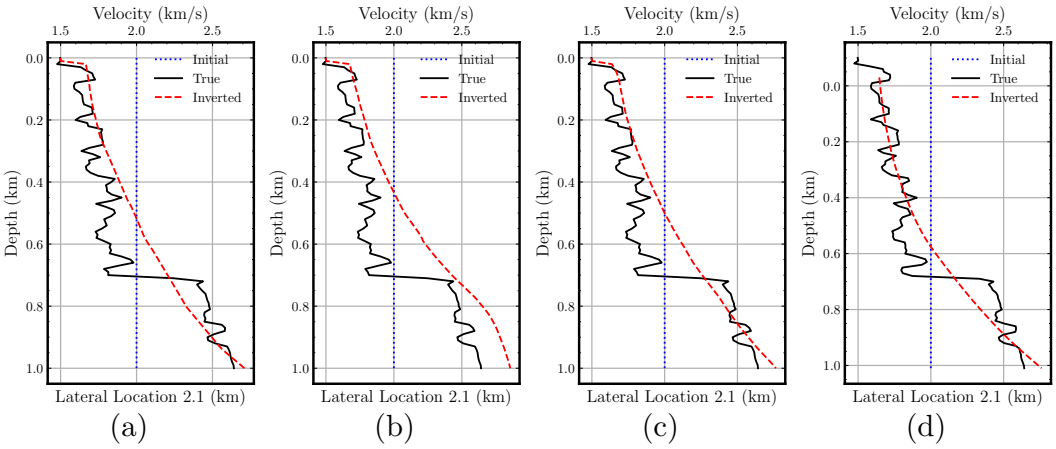

Figure 6. Vertical velocity profiles comparison for the 2D surface experiment for different scenarios in 5 (b-e). Profiles are extracted at 2.1 km lateral location.

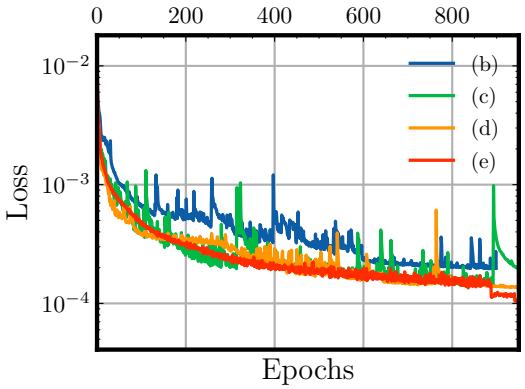

Figure 7. Loss curves for the training to obtain the inversion results shown in Figures 5(b-e).

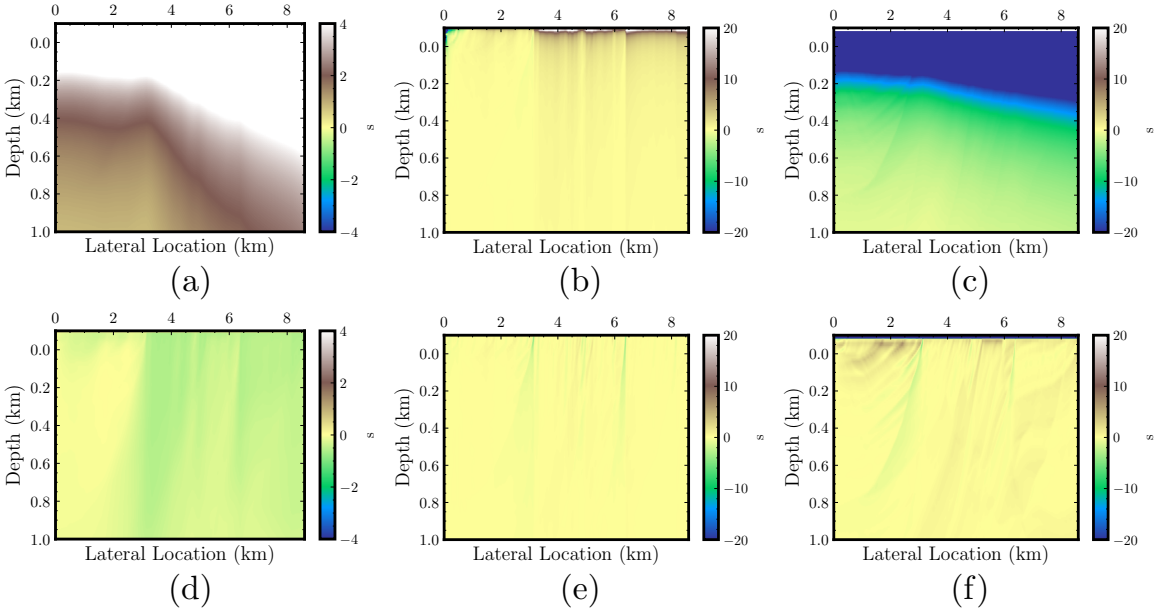

Figure 8. Factorization comparison. Multiplicative factorization, its lateral and vertical first-order derivatives extracted from the true $\hat{\tau}(\mathbf{x})$ are depicted in (a-c), respectively. The same configuration for the additive factorization approach are depicted in (d-f). The sections are extracted from the true $\hat{\tau}(\mathbf{x})$ field in Figure 5b.

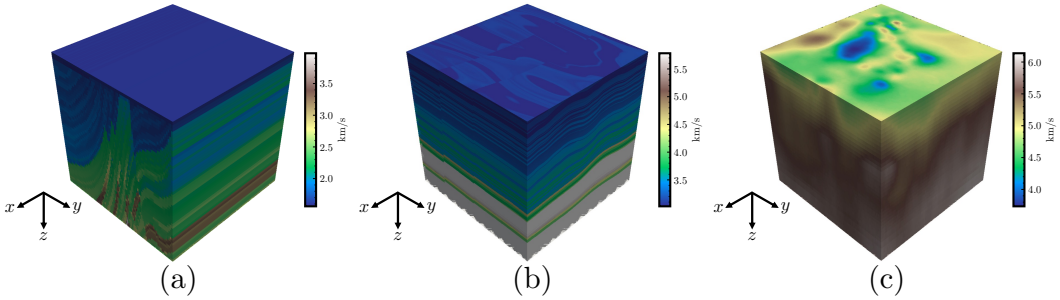

Figure 9. Three 3D velocity models used to simulate the travel time data. A 2.5D true velocity from a cropped Marmousi model (a). True 3D velocity models synthetic SEAM Phase II model (b) and derived from (White et al., 2021) (c).

multiplication in equation 2) to yield

$$T(\mathbf{x}) = \tau(\mathbf{x}) \cdot T_0(\mathbf{x}). \tag{13}$$

From our experiments, however, we find that for a hard-constraint PINN-based tomography, the additive factorization is more favorable than the multiplicative factorization. Figure 8, which represents a depth slice of the true $\hat{\tau}(\mathbf{x})$ for a surface tomography acquisition performed in the next section, exemplifies the point. These figures demonstrate that the additive factorization yields a much smoother true travel time factor ($\hat{\tau}(\mathbf{x})$) function. We can see that the $\hat{\tau}(\mathbf{x})$ values varies significantly between the depth near the source location ($z = 0.03$ km) and away from the source ($z = 1$ km) for the multiplicative factorization (Figure 8a), while the additive (Figure 8d) yields an almost constant function. The lateral and vertical (middle and right columns of Figure 8) further highlights the significant amplitude difference between the two approaches. Hence, with the known spectral bias feature of the NN (Rahaman et al., 2019), the travel time NN will favor predicting the additive (smoother) function over the multiplicative factorization function.

3.4 Surface Seismic Tomography in a 3D Medium

As mentioned earlier, the multi-term nature of previous PINN-based seismic tomography results in challenging training dynamics. Thus, previous PINN-based seismic tomography approaches, so far, only consider a 2D representation of the Earth. This section demonstrates the PINNs' ability to invert for a 3D velocity distribution by virtue of the more stable single-term loss function. We perform numerical simulation on three different 3D velocity models depicted in Figure 9.

3.4.1 2.5D Synthetic Velocity

We first consider a 2.5D velocity model. We duplicate the same cropped Marmousi model in the previous 2D subsection along the y axis depicted in Figure 9a. We use a 0.02 km grid spacing for the vertical and 0.1 km grid spacing for the lateral directions. The data are synthetically generated using a forward eikonal solver and recorded at $z = 0$ km. The sources are located near the surface and are regularly sampled with a lateral spacing of 1 km. We utilize a 12-layer with 12 neurons NN for the velocity NN and a 12-layer with 24 neurons for the travel time NN. All of these networks are initialized using the Glorot uniform initialization. These NNs use the ELU activation function and an Adam optimizer. The PINNs are trained for 1000 epochs using the generated 31,267,539 training samples ($\mathbf{x} = \{x, y, z\}$) with a 15,633 batch size. Throughout the subsequent experiments, we use an initial learning rate of 0.0005 and decrease its value by half ev-

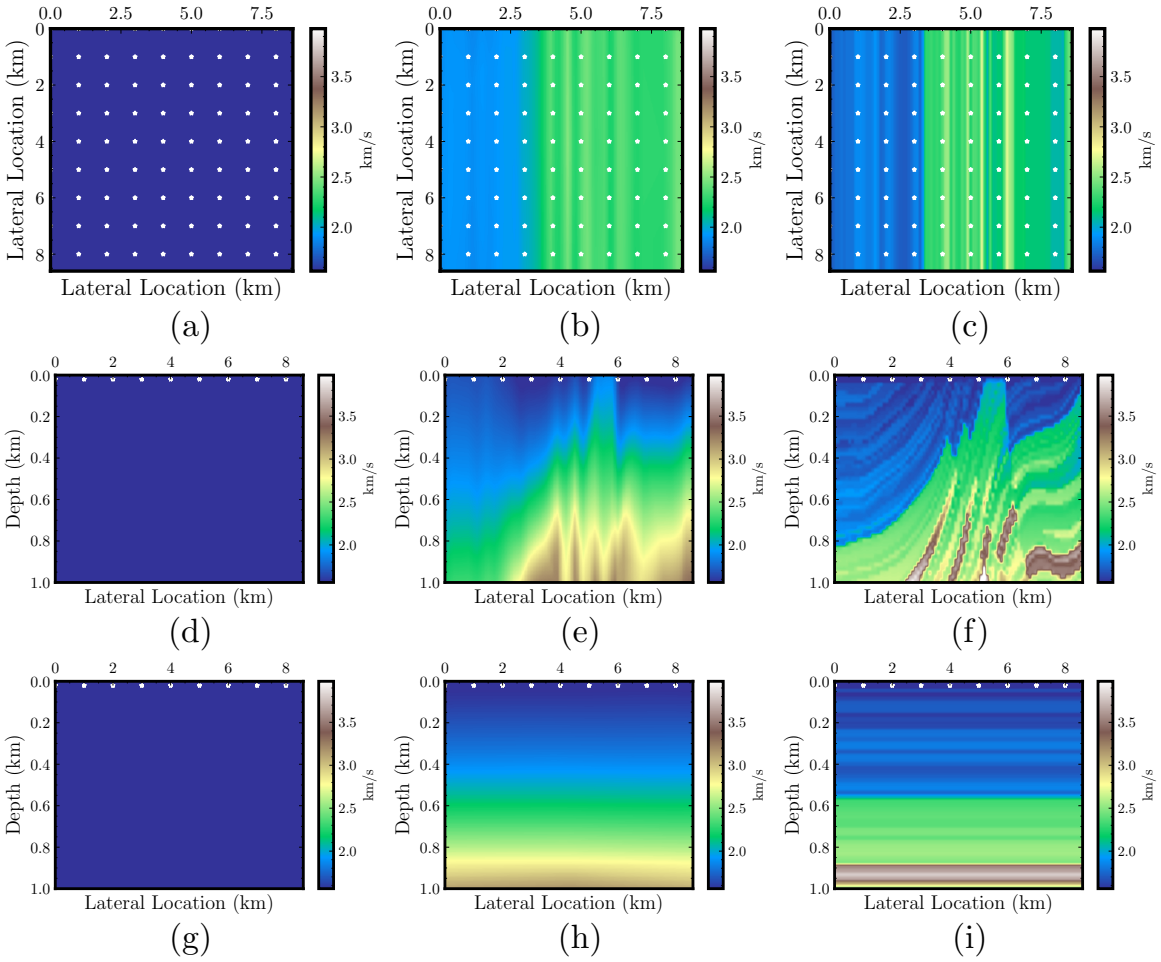

Figure 10. Velocity sections for the 2.5D cropped Marmousi model. The initial, inverted, and true velocity slices are depicted from the left to right columns. The top to bottom rows show the slices for xy, xz, yz planes on $z = 0.5$ km, $y = 3$ km, $x = 2$ km, respectively.

ery 200 training epochs. Further, we do not utilize the data NN for these regularly sampled measurements experiments.

Figure 10 offers a cross-sectional view of the inversion performance. We find a good agreement between the inverted velocity (middle column) and the true velocity model (right column). Figure 11 also demonstrates the accurate inverted velocity from a vertical profile view. Using the same computational domain and sample training points, we can further compare the 2D and 2.5D inversion process (Figure 12). By also comparing the inverted section between the 2D (Figure 5b) and 2.5D (Figure 10e), we observe a better and faster convergence of the 3D inversion. This is partially due to the fact that the rays cover a larger distance in the 2.5D model than the 2D case, and we have more illumination.

3.4.2 3D Synthetic Velocity

To investigate the framework's ability on a more realistic Earth model, we consider a synthetic experiment using the cropped 3D SEAM Phase II Arid model depicted in Figure 9b. We use a 0.0187 km grid spacing for the vertical and 0.075 km grid spacing

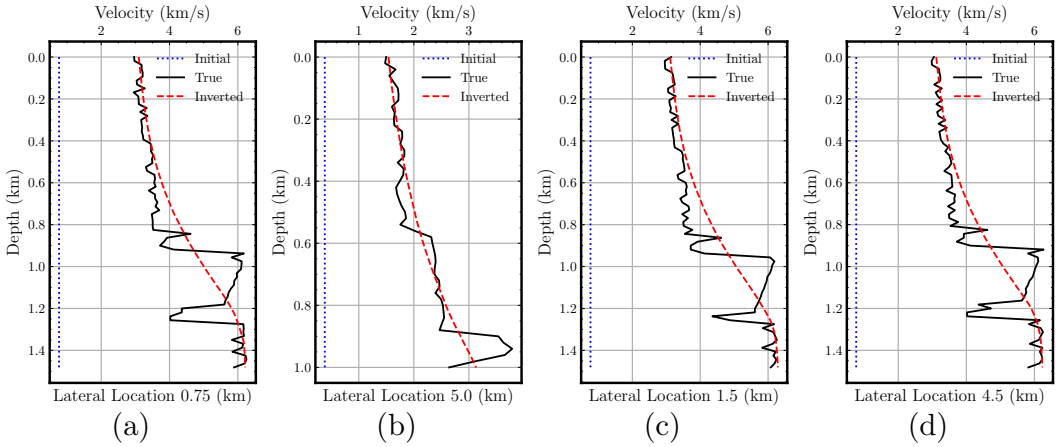

Figure 11. Vertical velocity profiles comparison for the 2.5D surface experiment at different lateral locations extracted from the inverted velocity in Figure 10.

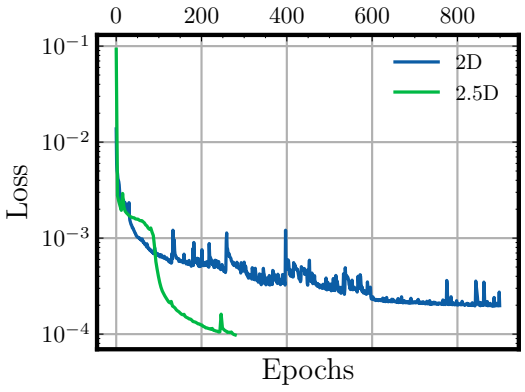

Figure 12. Loss curves comparison between the training for the inversion performed in Figure 5b and Figure 10e.

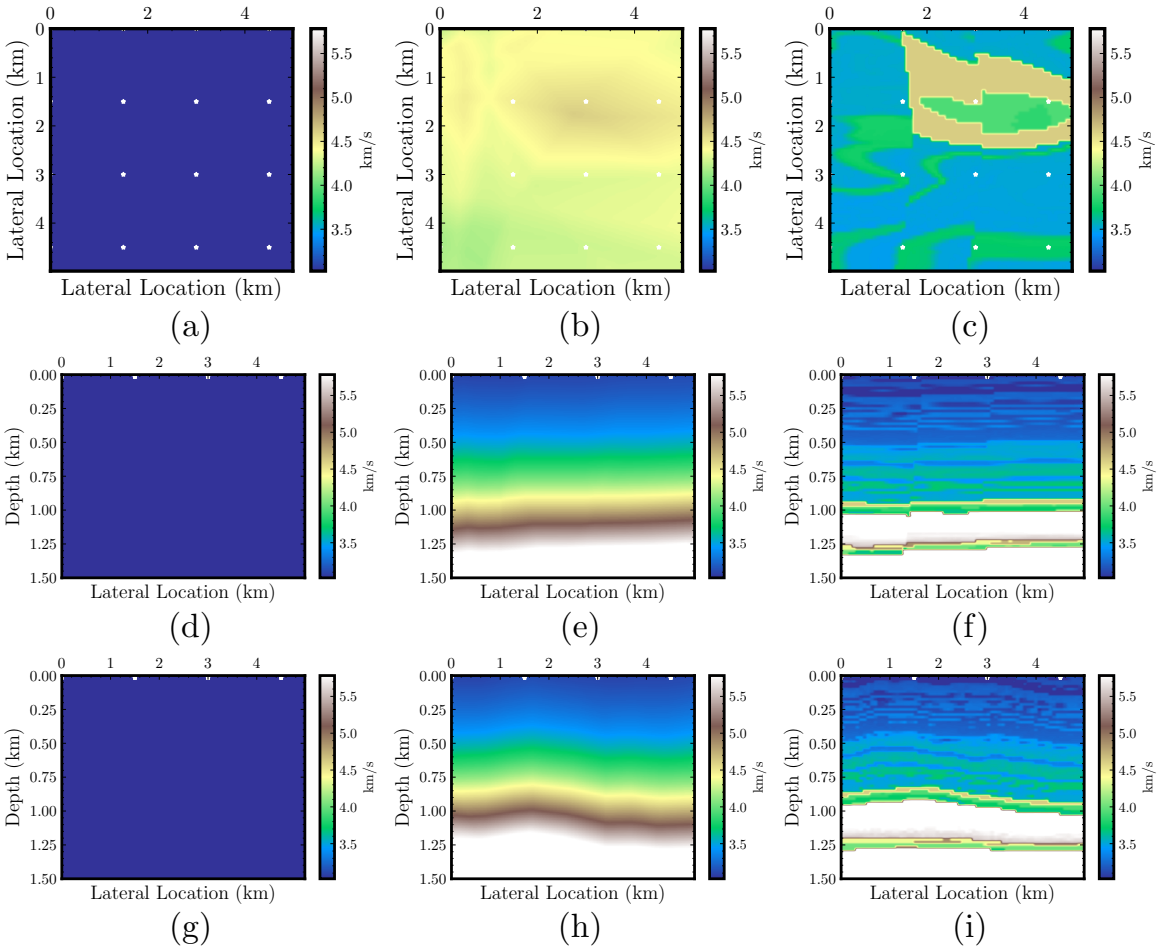

Figure 13. Velocity sections for the 3D SEAM Phase II Arid model. The initial, inverted, and true velocity slices are depicted from the left to right columns. The top to bottom row shows the slices for xy, xz, yz planes on $z = 0.84$ km, $y = 0.75$ km, $x = 2.25$ km, respectively.

for the lateral directions with a maximum lateral location and depth of 4.95 and 1.481 km, respectively. The data are synthetically generated using a forward eikonal solver and stored at $z = 0$ km. The sources are located near the surface and are regularly sampled with a lateral spacing of 1.5 km. The PINNs is trained for 1000 epochs using the generated 5,745,920 training samples ($\mathbf{x} = \{x, y, z\}$) with a 2,872 batch size. We utilize the same model complexity and training mechanism as in the previous 2.5D experiment.

The synthetic SEAM Phase II Arid is intended to mimic a challenging shallow crust (around the first 500 m) condition–typical to land exploration seismology cases. The right column of Figure 13 demonstrates the complex geological structures and lateral velocity variations of the model. The inverted velocity sections (middle column of Figure 13) show good agreement, in terms of the geological structures, with the input model. Both of the faulting systems (around 2 km lateral location in Figures 13e & f) as well as the anticlinal structure (around 2 km lateral location in Figures 13h & i) are recovered by the inversion process. The vertical profiles (Figure 14 also exemplifies the good inversion results as the inverted velocity captures the relatively constant velocity around the first 0.8 km depth and increases sharply at around 0.9 km depth.

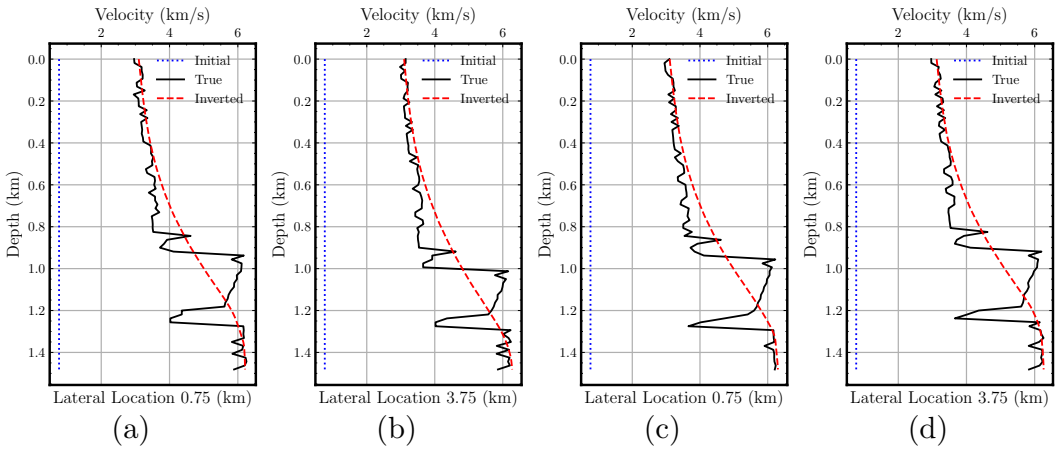

Figure 14. Vertical velocity profiles comparison for the 3D surface experiment at different lateral locations extracted from the inverted velocity in Figure 13.

3.4.3 3D Regional Velocity

Finally, to assess the performance of the framework's in dealing with a typical regional-scale eikonal tomography problem, we also perform a numerical experiment using inverted velocity provided from the study of (White et al., 2021) depicted in Figure 9c. It is worth mentioning that the purpose of this subsection is solely a further proof-of-concept that PINNs can as well be a viable option to deal with a typical regional-scale eikonal tomography. Compared to the previous 3D models, this velocity model contains more complex lateral velocity variation near the shallow crust. The data are synthetically generated using a forward eikonal solver and stored at $z = 0$ km. The sources are located near the surface and are regularly sampled with a lateral spacing of 32.637 km. The PINNs are trained for 1000 epochs using the generated 23,040,000 training samples ($\mathbf{x} = \{x, y, z\}$) with a 11,520 batch size. We utilize the same model complexity and training mechanism as in the previous 2.5D experiment.

Figure 16 illustrates the good agreement between the inverted and true (reference) velocity models. By comparing the inverted sections (middle column in Figure 16) and the true velocity (right column in Figure 16), we can see that the PINNs capture the complex shallow crust velocity variation. It is also worth noting that the inversion starts with a randomly initialized velocity model. This further demonstrates the stability of PINNs for eikonal tomography. Furthermore, we also find, at inference time (after the PINNs training), that we can instantly access the travel time field at a speed faster than a conventional eikonal solver (White et al., 2020). Running on the same Intel(R) Xeon(R) Gold 6230R CPU @ 2.10GHz, the trained PINNs model takes 3.68 minutes while the conventional eikonal solver takes 4.63 minutes.

4 Discussions

This work promotes a new formulation of a PINN-based eikonal tomography that relies on a single PDE residual term in its objective function. We reformulate the conventional eikonal tomography by replacing an eikonal solver with an NN that predicts the (factored) travel time field and another NN that substitutes the inversion kernel (e.g., adjoint-state method) to compute the velocity field. The minimization of the PINNs training will then be driven by obeying the PDE residual and a data misfit terms. In this case, the data measurement misfit is enforced using a hard constraint during the PINNs train-

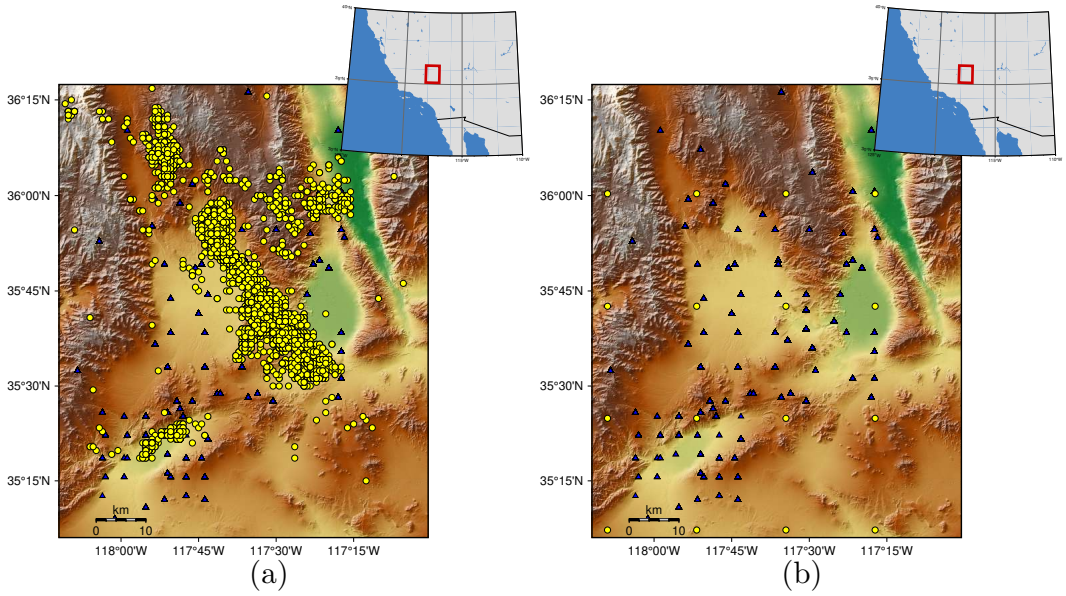

Figure 15. Map views of the events (yellow dots) and receivers (blue triangles) extracted from (White et al., 2021) (a) and the modified surface (z=0 km), regularly sampled, sources acquisition setup (b).

ing. Since lateral derivatives of the recorded travel time is needed, and to handle irregularly sampled measurements, we proposed the use of an NN to learn the measurements and inherit *tomographic* information from the source-receiver input location. Combining these three NNs as a framework, we found that hard-constrained measurement data significantly reduces the instability issues induced by the multi-term nature of the previous PINN-based eikonal tomography. These properties can be attributed to the natural inclusion of boundary conditions in the optimization problem. Thus, we alleviate a known pathology of the PINNs training in which a proper weighting scheme is needed to accommodate different loss function terms.

As a result, the single-term loss function improves the PINN-based eikonal tomography convergence. We found empirically that it can handle complex velocity distribution much better than the original PINN-based tomography formulation (Waheed et al., 2021a). It can even allow us to handle 3D measurements. Furthermore, the need for reliable measurement data can be partially achieved by a *tomographic* interpolation NN. We believe that by combining these components, PINN-based eikonal tomography can be a viable option for velocity model building at all scales. One benefit comes in the form that the trained velocity NN acts as a compressed representation of the velocity model (e.g., 3D velocity models). Another important feature is its flexibility in handling different acquisition scenarios. The method can still be applied without performing any coordinate transformation even for a topography-dependent recording surface. Moreover, as we have demonstrated, the travel time NN performs more efficiently than a conventional eikonal solver.

We have shown in the context of eikonal tomography that we can provide a stable inversion by reducing the number of loss terms in the training process. However, PINN-based approaches still require further development as an emerging option for further seismic tomography applications. Fortunately, the method presented here is orthogonal with previous attempts that view improper sampling as the major source of poor performance of PINNs. In other words, further development can accommodate a better sampling strat-

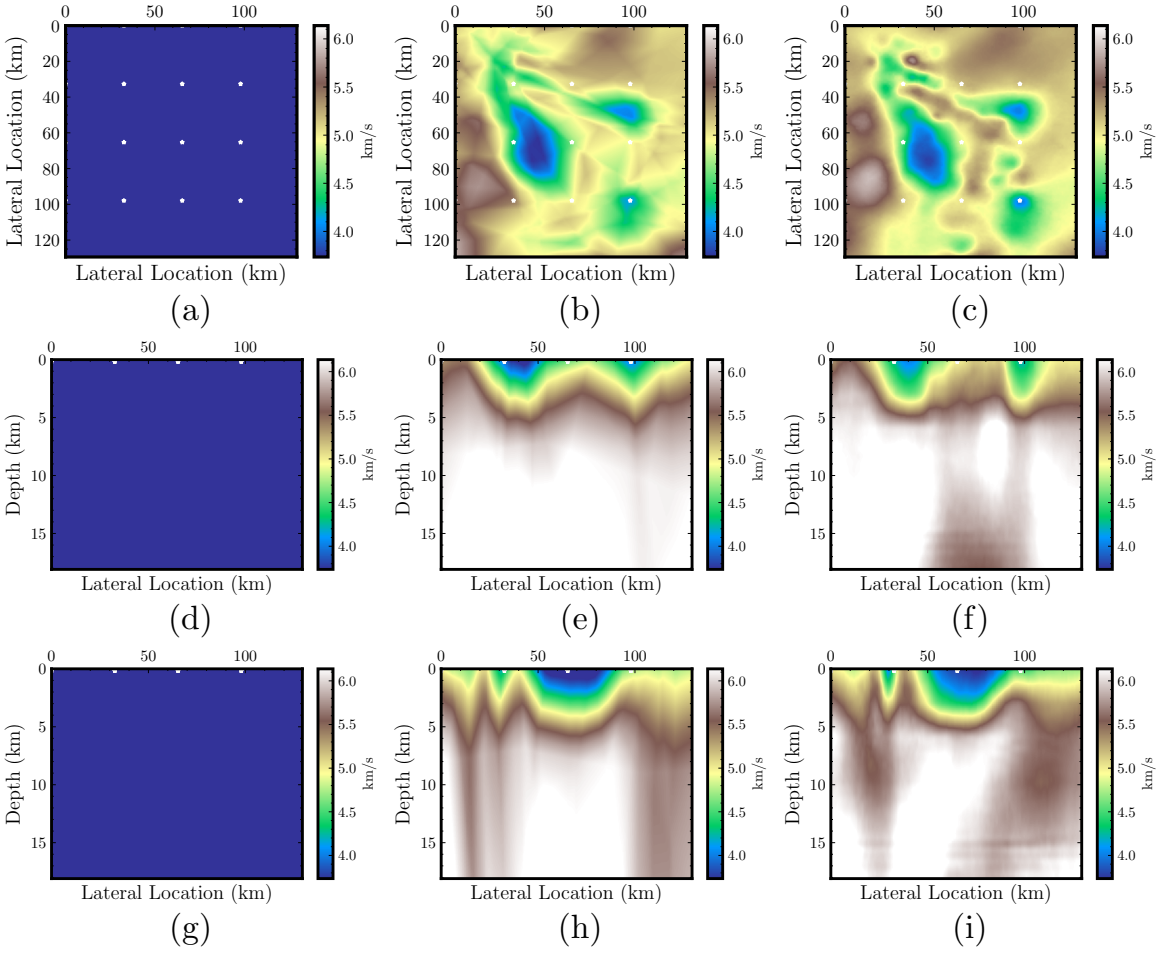

Figure 16. Velocity sections for the 3D Earth model from (White et al., 2021). The initial, inverted, and true velocity slices are depicted from the left to right columns. The top to bottom row shows the depth, latitude, and longitude sections on 0.91 km, 35.6° N, 117.8° W, respectively.

egy that, for example, obeys causality (Wang et al., 2022) and accounts for samples that have higher residual errors (Yu et al., 2022). We have also demonstrated the capability of PINNs to handle complex and realistic velocity models, the remaining challenge now is on how we further accelerate the training convergence, especially for large data measurements. Furthermore, for industrial-size 3D field data measurements, distributed multi-GPU training is inevitable. As in such a scenario, a single-GPU setup will run into the issue of out-of-memory problems. We anticipate that a multi-GPU setup with a mixed-even half-precision (to accelerate the convergence) can be made possible courtesy of the stable single-term loss function. Therefore, such a method can be a starting way forward to utilize PINNs for large-scale eikonal tomography.

5 Conclusions

We proposed a new framework for eikonal tomography, which is more suitable for a PINN-based tomographic workflow. We train, simultaneously from randomly initialized weights, two NNs to invert for the phase velocity and travel time field. We also suggest an optional data interpolation NN that is embedded with tomographic information. The new framework allows us to rely on a stable single-term loss function. The stability of the PINNs inversion can also be attributed to the additive factorization used to decompose the travel time field. We found that this factorization will result in a more stable PINN-based tomography compared to the multiplicative version. With rigorous numerical experimentation, we have demonstrated that it leads to a robust and flexible inversion framework that handles 3D Earth's model. Our new framework opens up new possibilities in dealing with sparse data measurements, that is typical in global seismological applications.

6 Data Availability Statement

Codes and data needed to reproduce the results presented here will be made available at https://github.com/hatsyim/HCPINNsEikonal.

References

Aki, K., Christoffersson, A., & Husebye, E. S. (1977). Determination of the three-dimensional seismic structure of the lithosphere. *Journal of Geophysical Research*, *82*(2), 277–296.

Alkhalifah, T. (2002, 7). Traveltime computation with the linearized eikonal equation for anisotropic media. *Geophysical Prospecting*, *50*, 373-382. Retrieved from `https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-2478.2002.00322.xhttps://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2478.2002.00322.xhttps://onlinelibrary.wiley.com/doi/10.1046/j.1365-2478.2002.00322.x` doi: 10.1046/J.1365-2478.2002.00322.X

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. (2018). Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research*, *18*(153), 1-43. Retrieved from `http://jmlr.org/papers/v18/17-468.html`

Bergman, B., Tryggvason, A., & Juhlin, C. (2004). High-resolution seismic traveltime tomography incorporating static corrections applied to a till-covered bedrock environment. *Geophysics*, *69*(4), 1082–1090.

Bording, R. P., Gersztenkorn, A., Lines, L. R., Scales, J. A., & Treitel, S. (1987). Applications of seismic travel-time tomography. *Geophysical Journal International*, *90*(2), 285–303.

Červený, V. (2000). *Seismic ray method*. Cambridge University Press.

Chen, Y., de Ridder, S. A., Rost, S., Guo, Z., Wu, X., & Chen, Y. (2022). Eikonal tomography with physics-informed neural networks: Rayleigh wave phase ve-

locity in the northeastern margin of the tibetan plateau. *Geophysical Research Letters*, *49*(21), e2022GL099053.

Dessa, J.-X., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., Uhira, K., & Kaneda, Y. (2004). Deep seismic imaging of the eastern nankai trough, japan, from multifold ocean bottom seismometer data by combined travel time tomography and prestack depth migration. *Journal of Geophysical Research: Solid Earth*, *109*(B2).

Dziewonski, A. M., Hager, B. H., & O'Connell, R. J. (1977). Large-scale heterogeneities in the lower mantle. *Journal of Geophysical Research*, *82*(2), 239–255.

Fang, H., White, M. C., Lu, Y., & Ben-Zion, Y. (2022). Seismic traveltime tomography of southern california using poisson-voronoi cells and 20 years of data. *Journal of Geophysical Research: Solid Earth*, *127*(5), e2021JB023307.

Fang, H., Zhang, H., Yao, H., Allam, A., Zigone, D., Ben-Zion, Y., ... van der Hilst, R. D. (2016). A new algorithm for three-dimensional joint inversion of body wave and surface wave data and its application to the southern california plate boundary region. *Journal of Geophysical Research: Solid Earth*, *121*(5), 3557–3569.

Fichtner, A. (2010). *Full seismic waveform modelling and inversion*. Springer Science & Business Media.

Fomel, S., Luo, S., & Zhao, H. (2009, 9). Fast sweeping method for the factored eikonal equation. *Journal of Computational Physics*, *228*, 6440-6455. doi: 10.1016/j.jcp.2009.05.029

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256).

Hauksson, E., Yang, W., & Shearer, P. M. (2012). Waveform relocated earthquake catalog for southern california (1981 to june 2011). *Bulletin of the Seismological Society of America*, *102*(5), 2239–2244.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*, 359-366.

Julian, B., Gubbins, D., et al. (1977). Three-dimensional seismic ray tracing. *Journal of Geophysics*, *43*(1), 95–113.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Klein, F. (2002). User's guide to hypoinverse-2000, a fortran program to solve for earthquake locations and magnitudes. open file report 02-171. *US Geological Survey*, 1–123.

Leung, S., & Qian, J. (2006). An adjoint state method for three-dimensional transmission traveltime tomography using first-arrivals.

Lin, F.-C., & Ritzwoller, M. H. (2011). Helmholtz surface wave tomography for isotropic and azimuthally anisotropic structure. *Geophysical Journal International*, *186*(3), 1104–1120.

Lin, F.-C., Ritzwoller, M. H., & Snieder, R. (2009). Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array. *Geophysical Journal International*, *177*(3), 1091–1110.

Lin, F.-C., Ritzwoller, M. H., Townend, J., Bannister, S., & Savage, M. K. (2007). Ambient noise rayleigh wave tomography of new zealand. *Geophysical Journal International*, *170*(2), 649–666.

Marsden, D. (1993). Static corrections—a review, part 1. *The leading edge*, *12*(1), 43–49.

McClenny, L., & Braga-Neto, U. (2020). Self-adaptive physics-informed neural networks using a soft attention mechanism. *arXiv preprint arXiv:2009.04544*.

Obrebski, M., Allen, R. M., Zhang, F., Pan, J., Wu, Q., & Hung, S.-H. (2012). Shear wave tomography of china using joint inversion of body and surface wave

constraints. *Journal of geophysical research: solid earth*, *117*(B1).

Plessix, R.-E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, *167*(2), 495–503.

Qin, F., Luo, Y., Olsen, K. B., Cai, W., & Schuster, G. T. (1992). Finite-difference solution of the eikonal equation along expanding wavefronts. *Geophysics*, *57*(3), 478–487.

Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., ... Courville, A. (2019). On the spectral bias of neural networks. In *International conference on machine learning* (pp. 5301–5310).

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019, 2). 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. doi: 10.1016/j.jcp.2018.10.045

Rawlinson, N., & Fishwick, S. (2012). Seismic structure of the southeast australian lithosphere from surface and body wave tomography. *Tectonophysics*, *572*, 111–122.

Rawlinson, N., Hauser, J., & Sambridge, M. (2008, 1). Seismic ray tracing and wavefront tracking in laterally heterogeneous media. *Advances in Geophysics*, *49*, 203-273. doi: 10.1016/S0065-2687(07)49003-3

Rawlinson, N., & Sambridge, M. (2004). Wave front evolution in strongly heterogeneous layered media using the fast marching method. *Geophysical Journal International*, *156*(3), 631–647.

Schiassi, E., Leake, C., Florio, M. D., Johnston, H., Furfaro, R., & Mortari, D. (2020, 5). Extreme theory of functional connections: A physics-informed neural network method for solving parametric differential equations. *arXiv preprint arXiv:2005.10632*. Retrieved from `https://arxiv.org/abs/2005.10632v1` doi: 10.48550/arxiv.2005.10632

Sethian, J. A. (1996). A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, *93*(4), 1591–1595. Retrieved from `https://www.pnas.org/content/93/4/1591` doi: 10.1073/pnas.93.4.1591

Simmons, N. A., Myers, S. C., Johannesson, G., & Matzel, E. (2012). Llnl-g3dv3: Global p wave tomography model for improved regional and teleseismic travel time prediction. *Journal of Geophysical Research: Solid Earth*, *117*(B10).

Simmons, N. A., Myers, S. C., Morency, C., Chiang, A., & Knapp, D. R. (2021). Spiral: a multiresolution global tomography model of seismic wave speeds and radial anisotropy variations in the crust and mantle. *Geophysical Journal International*, *227*(2), 1366–1391.

Smith, J. D., Azizzadenesheli, K., & Ross, Z. E. (2021). Eikonet: Solving the eikonal equation with deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(12), 10685-10696. doi: 10.1109/TGRS.2020.3039165

Taillandier, C., Noble, M., Chauris, H., & Calandra, H. (2009). First-arrival traveltime tomography based on the adjoint-state method. *Geophysics*, *74*. Retrieved from `http://library.seg.org/page/policies/terms` doi: 10.1190/1.3250266

Taufik, M. H., Waheed, U., & Alkhalifah, T. (2022). Upwind, no more: Flexible traveltime solutions using physics-informed neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-12. doi: 10.1109/TGRS.2022.3218754

Tavakoli, F. B., Operto, S., Ribodetti, A., & Virieux, J. (2017). Slope tomography based on eikonal solvers and the adjoint-state method. *Geophysical Journal International*, *209*(3), 1629–1647.

Thurber, C. H. (1983). Earthquake locations and three-dimensional crustal structure in the coyote lake area, central california. *Journal of Geophysical Research:*

Solid Earth, *88*(B10), 8226–8236.

Um, J., & Thurber, C. (1987). A fast algorithm for two-point seismic ray tracing. *Bulletin of the Seismological Society of America*, *77*(3), 972–986.

Vidale, J. (1988, 12). Finite-difference calculation of travel times. *Bulletin of the Seismological Society of America*, *78*(6), 2062-2076. Retrieved from `https://doi.org/10.1785/BSSA0780062062` doi: 10.1785/BSSA0780062062

Virieux, J., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, *74*(6), WCC1–WCC26.

Waheed, U., Alkhalifah, T., Haghighat, E., Song, C., & Virieux, J. (2021a, 4). Pintomo: Seismic tomography using physics-informed neural networks. *arXiv preprint arXiv:2104.01588*. doi: 10.48550/arxiv.2104.01588

Waheed, U., Haghighat, E., Alkhalifah, T., Song, C., & Hao, Q. (2021b, 10). Pineik: Eikonal solution using physics-informed neural networks. *Computers Geosciences*, *155*, 104833. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S009830042100131X` doi: 10.1016/j.cageo.2021.104833

Wang, S., Sankaran, S., & Perdikaris, P. (2022). Respecting causality is all you need for training physics-informed neural networks. *arXiv preprint arXiv:2203.07404*.

White, M. C., Fang, H., Catchings, R. D., Goldman, M. R., Steidl, J. H., & Ben-Zion, Y. (2021). Detailed traveltime tomography and seismic catalogue around the 2019 m w7. 1 ridgecrest, california, earthquake using dense rapid-response seismic data. *Geophysical Journal International*, *227*(1), 204–227.

White, M. C., Fang, H., Nakata, N., & Ben-Zion, Y. (2020). Pykonal: a python package for solving the eikonal equation in spherical and cartesian coordinates using the fast marching method. *Seismological Research Letters*, *91*(4), 2378–2389.

Yu, J., Lu, L., Meng, X., & Karniadakis, G. E. (2022). Gradient-enhanced physics-informed neural networks for forward and inverse pde problems. *Computer Methods in Applied Mechanics and Engineering*, *393*, 114823.

Zelt, C. A., Azaria, A., & Levander, A. (2006). 3d seismic refraction traveltime tomography at a groundwater contamination site. *Geophysics*, *71*(5), H67–H78.

Zelt, C. A., & Barton, P. J. (1998). Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the faeroe basin. *Journal of Geophysical Research: Solid Earth*, *103*(B4), 7187–7210.

Zhang, H., & Thurber, C. H. (2003). Double-difference tomography: The method and its application to the hayward fault, california. *Bulletin of the Seismological Society of America*, *93*(5), 1875–1889.

Zhang, J., & Toksöz, M. N. (1998). Nonlinear refraction traveltime tomography. *Geophysics*, *63*(5), 1726–1737.

Zhao, H. (2005). A fast sweeping method for eikonal equations. *Math. Comput.*, *74*, 603-627.

A Stable Neural Network-Based Eikonal Tomography using Hard-Constrained Measurements

Mohammad H. Taufik [1*], Tariq Alkhalifah[1], and Umair bin Waheed [2]

[1]Physical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, 23955, Saudi Arabia
[2]Department of Geosciences, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia

Key Points:

- We develop a novel implementation of eikonal tomography via physics-informed neural networks (PINNs), in which we embed the measurements (data) as hard constraints into the eikonal equation.
- We demonstrate the efficiency and flexibility of the new formulation for different use cases (no modifications to the eikonal), which marks a significant improvement from the conventional finite-difference based methods.
- The new formulation yields a single, stable, objective function removing the burden of figuring out an appropriate weighting for the loss terms of the previous PINN-based approaches. The algorithm also demonstrates its stability in handling complex 3D velocity distributions.

*Physical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, 23955, Saudi Arabia

Corresponding author: Mohammad Taufik, mohammad.taufik@kaust.edu.sa

Abstract

Eikonal tomography, or travel time inversion, has been one of the primary seismological tools for decades and has been used to understand Earth's properties and dynamic processes. At the heart of the inversion process is the need for an accurate, and preferably flexible, eikonal solver to compute the travel time field. Most of the conventional eikonal solvers, however, suffer from first-order convergence errors and difficulties in dealing with irregular computational grids. Physics-informed neural networks (PINNs) have been introduced to tackle these problems and have shown great success in addressing those challenges. Nevertheless, these approaches still suffer from slow convergence and unstable training dynamics due to the multi-term nature of the loss function. To improve on this, we propose a new formulation for the isotropic eikonal equation, which imposes boundary conditions as hard constraints. We employ the theory of functional connections to the eikonal tomography problem, which allows for the utilization of a single loss term for training the PINN model. Through rigorous numerical tests, its efficiency, stability, and flexibility in tackling a variety of cases, including topography-dependent and 3D models, are attested, thus providing an efficient and stable PINN-based eikonal tomography.

Plain Language Summary

Machine learning techniques have proven to be beneficial in various scientific and engineering applications. One of these techniques, termed as physics-informed neural networks (PINNs), attempts to solve parametric partial differential equations (PDEs) by imposing the necessary conditions and PDE residuals into the objective function being optimized. The multi-component nature of the PINNs' objective function, however, often renders the training to be unstable. Here, we attempt to circumvent this pathology by reformulating the PDE (eikonal equation) such that the necessary conditions (travel time measurements) are naturally included in the PDE. Thus, the objective function yields a single term that stabilizes the training process. We report that the proposed formulation is also flexible to be incorporated into different eikonal tomography applications from 2D to 3D Earth's model seamlessly.

1 Introduction

Seismic tomography (inversion) has been and still is one of the primary seismological tools for understanding the properties and description of the dynamic processes inside the Earth. Pioneered by the work of Dziewonski et al. (1977) and Aki et al. (1977), eikonal tomography facilitates the tomographic process by inverting the measured travel time information of the seismic waves. Although limited by the underlying ray theoretic high-frequency asymptotic assumption, decades of development and utilization justify the usability of the method in a wide range of seismological applications. In global seismology, eikonal tomography is regularly used to locate earthquakes (Thurber, 1983; Klein, 2002; Hauksson et al., 2012), image the Earth's interior via body waves (H. Zhang & Thurber, 2003), surfaces waves (Lin et al., 2007, 2009; Lin & Ritzwoller, 2011), joint surface-body waves tomography (Rawlinson & Fishwick, 2012; Obrebski et al., 2012; Fang et al., 2016), and deriving global velocity models (Simmons et al., 2012, 2021). On the exploration front, it is heavily involved in the imaging of shallow crust (J. Zhang & Toksöz, 1998; Zelt & Barton, 1998; Zelt et al., 2006), velocity model building for migration (Marsden, 1993; Dessa et al., 2004; Bergman et al., 2004) and full-waveform inversion (Virieux & Operto, 2009; Fichtner, 2010; Tavakoli et al., 2017), and reservoir monitoring and characterization (Bording et al., 1987).

At the heart of the travel time inversion process lies the eikonal equation. The eikonal equation is a first-order nonlinear partial differential equation (PDE) that can be derived, for example, from the infinite-frequency assumption of the wave equation. The tomographic algorithms can be performed by either performing linearization to the inversion

operator (e.g., via ray theory (Červený, 2000) or via the eikonal (Alkhalifah, 2002)) or using the adjoint-state method (Leung & Qian, 2006; Plessix, 2006; Taillandier et al., 2009). The conventional linearization approach gives rise to the need of computing (and storing) the Fréchet matrix, and its inverse, which is a challenge for 3D dense measurements. The adjoint-state approach avoids the calculation of such a matrix and has shown promising results over the years. Shared across these algorithms is the need for an efficient and preferably flexible way to solve the eikonal equation. To achieve this, conventional algorithms resort to either the ray-based approaches (Červený, 2000; Julian et al., 1977; Um & Thurber, 1987) or the grid-based approaches (Vidale, 1988; Qin et al., 1992; Sethian, 1996; Rawlinson & Sambridge, 2004; Zhao, 2005). However, most of these algorithms still suffer from the so-called first-order convergence error and sharp velocity changes; in these scenarios, the algorithms may fail to converge (Rawlinson et al., 2008). More importantly, a significant modification (e.g., coordinate transformation) needs to be incorporated when dealing with irregular surface topography.

On another front, employing physics-informed neural networks (PINNs) (Raissi et al., 2019) as a replacement to the conventional eikonal solver has shown promising results in addressing these issues. Smith et al. (2021) and Waheed et al. (2021b) demonstrated the more accurate travel time fields produced by the PINN-based solver for travel time modeling and identify the mesh-independent nature of such solvers. Taufik et al. (2022) demonstrated the nonlinear interpolation ability of the PINN-based solver in dealing with non-uniformly sampled data. Moreover, Waheed et al. (2021a) proposed a framework for treating the ill-posed body wave tomography problems by simultaneously utilizing two neural networks (NNs) to invert for the travel time and velocity fields. Chen et al. (2022) extended the framework to invert for a surface wave tomography application.

These successful cases can partially be explained by a well-known nonlinear interpolation ability of the NNs (Hornik et al., 1989) as well as the use of travel time factorization approaches. Travel time factorization is introduced to overcome the so-called source-singularity problem (Fomel et al., 2009). The problem statement then becomes, given an initial (background) travel time field, the NN is trained to predict a variable that maps the initial to the actual travel time field. Therefore, the factorization implicitly drives the PINNs' loss function to include at least two terms, namely the PDE residual and the boundary-related condition loss. For travel time tomography, both the boundary condition and data mismatch need to be explicitly imposed during the training process of PINNs. Hence, on top of the already challenging training dynamics, the multi-term nature of the PINNs' training under this regime often induces more instability due to the need to properly balance the loss terms with respect to their weight and the number of samples.

Several previous works attribute the PINNs' instability to improper loss function sampling. Wang et al. (2022) reformulated the PINNs' objective functions by explicitly obeying the physical causality. McClenny and Braga-Neto (2020) proposed the use of an efficient weighting scheme to the multi-term objective function. Based on this idea, (Yu et al., 2022) utilized an effective sampling based on the residual values. In this work, orthogonal to the aforementioned approaches, we develop a new formulation for the isotropic eikonal equation by imposing the boundary conditions as hard constraints (HC). We implement the theory of functional connections (TFC) (Schiassi et al., 2020) into the eikonal tomography problem, which admits a single loss term for training the PINNs-based model. Trained in this fashion, not only does our formulation still inherit all the known favorable properties of PINNs, but it also makes the training more robust and efficient. These properties are the result of the natural inclusion of boundary conditions in the optimization problem. Combining this with a data interpolation NN makes the proposed scheme handle sparse and irregularly sampled measurements accurately.

Thus, the contributions of our work can be summarized as follows,

1. Introduce a novel PINN-based eikonal tomography framework that inverts for the velocity and travel time fields with a single loss term for training the PINNs model.
2. Derive a new isotropic eikonal equation with hard-constrained data measurements.
3. Propose the use of a nonlinear data interpolation scheme to handle sparse and irregularly sampled measurements.
4. Analyze different travel time factorization approaches that best suit PINN-based eikonal tomography.
5. Provide, to the best of our knowledge, the first evidence of the use of PINN-based eikonal tomography to handle regional-scale 3D data.

In the following, we begin by providing the necessary theoretical background to derive the proposed framework, followed by several numerical experiments to showcase the flexibility and efficiency of the framework in different acquisition scenarios.

2 Methodology

In this section, we will first present the governing eikonal equation in its factorized form. We will then discuss how to embed the measured travel time data as hard constraints in the eikonal equation. Finally, we share the rationale for using a neural network to perform nonlinear data interpolation.

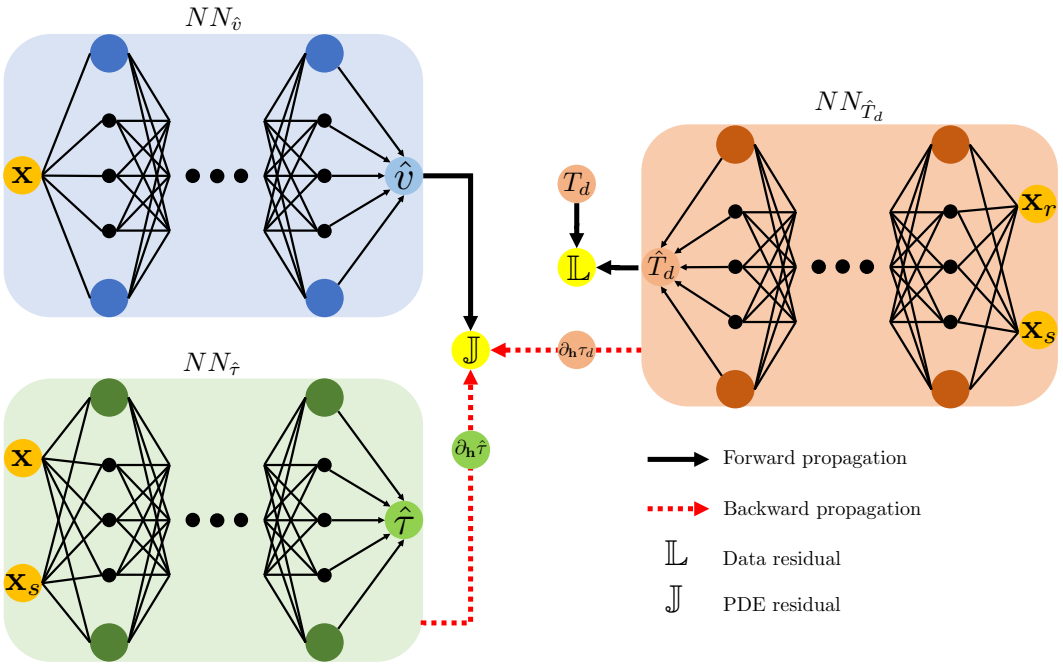

Figure 1. The proposed workflow for our PINN-based eikonal tomography problem.

2.1 The Isotropic Eikonal and Its Factorization

The eikonal equation for an isotropic medium can be written in the form of (Červený, 2000):

$$|\nabla T(\mathbf{x})|^2 = \frac{1}{v^2(\mathbf{x})}, \tag{1}$$

where T denotes the travel time field and v denotes the medium phase velocity, both as a function of position vector $\mathbf{x} = \{x, y, z\} \in \mathbb{R}^3$ for a three-dimensional Cartesian sys-

tem. To mitigate the source singularity problem, the travel time field can also be split into an additive form, which yields the so-called factored eikonal equation given by:

$$T(\mathbf{x}) = \tau(\mathbf{x}) + T_0(\mathbf{x}), \tag{2}$$

where a scalar τ is introduced to map the background travel time T_0 to the actual travel time T. Specifically, the background travel time is simply defined as:

$$T_0(\mathbf{x}) = \frac{\sqrt{(\mathbf{x} - \mathbf{x}_s)^2}}{v_0(\mathbf{x})}, \tag{3}$$

which is given by the distance from the source location $\{\mathbf{x}_s\}$ divided over some background velocity, e.g., a constant velocity (v_0).

Using equation 2 to represent the travel time field T, we have,

$$\nabla T(\mathbf{x}) = \nabla\tau(\mathbf{x}) + \nabla T_0(\mathbf{x}). \tag{4}$$

2.2 Incorporating Data as Hard Constraint

To incorporate the measured travel time data in the eikonal equation as a hard constraint, we use the theory of connection functions (Schiassi et al., 2020). For this derivation, we first consider a typical surface tomography experiment in which the data measurements T_d take place along a constant depth surface at z_r. In this case, we suggest the following representation of the travel time field:

$$T(\mathbf{x}) = \zeta(z)\hat{\tau}\left(\mathbf{x}\right) + \tau_d\left(x, y\right) + T_0\left(\mathbf{x}\right), \tag{5}$$

where

$$\zeta(z) = z - z_r, \tag{6}$$

and

$$\tau_d(x, y) = T_d(x, y) - T_0(x, y, z = z_r). \tag{7}$$

The term $\hat{\tau}(\mathbf{x})$ is parameterized by a neural network (NN) functional. Alternatively, for a typical cross-hole measurement, where the source and receiver locations are fixed at a certain lateral location, we can easily adapt equation 5 using

$$\zeta(x) = x - x_r. \tag{8}$$

More generally, we can even further impose a topography-dependent recording surface in which the z_r in equation 6 is a function of x (for a surface tomography problem). Compared to the original factored eikonal equation, here the new factor $\hat{\tau}$ is not guaranteed to be positive everywhere in the domain of interest.

To this end, the overall workflow of our proposed PINN-based tomography is demonstrated in Figure 1. The core of the inversion is shown by the left two NNs while the right NN corresponds to the optional data fitting/interpolation NN ($NN_{\hat{T}_d}$) trained prior to the main inversion (training) process. The PINNs tomography model consist of two NNs, trained simultaneously from randomly initialized weights, to predict the velocity ($NN_{\hat{v}}$) and the travel time ($NN_{\hat{\tau}}$). The inputs to the travel time NN function are the location of the source, $\{\mathbf{x}_s\}$ and the position in space $\{\mathbf{x}\}$, whereas only $\{\mathbf{x}\}$ are inputs to the velocity NN function. Using automatic differentiation (AD) (Baydin et al., 2018), we can compute the lateral travel time derivatives ($\partial_{\mathbf{h}}\hat{\tau}$). The data NN function of the source and sensor locations is trained in a supervised fashion prior to the PINNs' training using a loss function, $\mathbb{L}$, that measures the misfit between the predicted data and the measured data T_d. The data network takes as input the location of the source $\{\mathbf{x}_s\}$ and receiver $\{\mathbf{x}_r\}$. Once trained, the data travel time NN will not only provide us with the ability to handle sparse measurements, it also admits the gradient of the travel time data

through AD ($\partial_{\mathbf{h}}\hat{\tau}_d$), which is needed for the new eikonal as we will see. We can compute the gradient either prior or during the PINNs training.

Using equation 4, the travel time gradient components of the eikonal can be formulated as:

$$\frac{\partial T\left(\mathbf{x}\right)}{\partial x} = \zeta(z)\frac{\partial \hat{\tau}\left(\mathbf{x}\right)}{\partial x} + \frac{\partial \tau_d\left(x,y\right)}{\partial x} + \frac{\partial T_0\left(\mathbf{x}\right)}{\partial x}, \tag{9}$$

$$\frac{\partial T\left(\mathbf{x}\right)}{\partial y} = \zeta(z)\frac{\partial \hat{\tau}\left(\mathbf{x}\right)}{\partial y} + \frac{\partial \tau_d\left(x,y\right)}{\partial y} + \frac{\partial T_0\left(\mathbf{x}\right)}{\partial y}, \tag{10}$$

and

$$\frac{\partial T\left(\mathbf{x}\right)}{\partial z} = \zeta(z)\frac{\partial \hat{\tau}\left(\mathbf{x}\right)}{\partial z} + \frac{\partial \zeta(z)}{\partial z}\hat{\tau}\left(\mathbf{x}\right) + \frac{\partial T_0\left(\mathbf{x}\right)}{\partial z}. \tag{11}$$

Hence, the proposed loss function for the PINN, $\mathbb{J}$, can be constructed by plugging in the gradients from the travel time network (equations 9 to 11) and the velocity $\hat{v}$ from the velocity network into equation 1.

$$\mathbb{J} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial T\left(\mathbf{x}_i\right)}{\partial x}^2 + \frac{\partial T\left(\mathbf{x}_i\right)}{\partial y}^2 + \frac{\partial T\left(\mathbf{x}_i\right)}{\partial z}^2 - \frac{1}{\hat{v}^2(\mathbf{x}_i)}\right). \tag{12}$$

2.3 Non-linear Data Interpolation

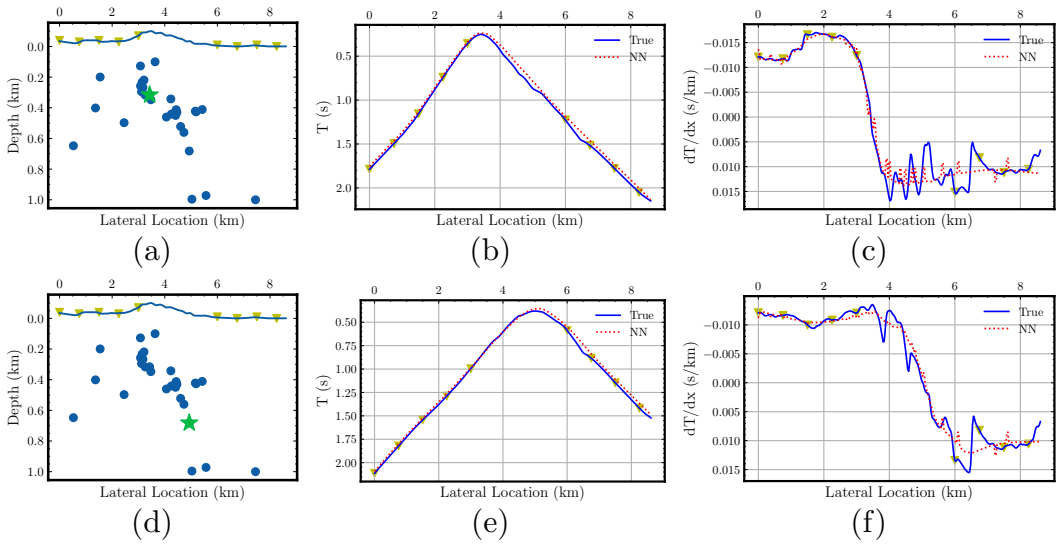

Figure 2. The data NN interpolation demonstration. The acquisition setup (left column, a and d). Travel time data and its horizontal derivatives (middle and right columns). The events used in the training are depicted by the blue dots, while results in the middle and right columns correspond to the green star event, and the receivers are denoted by the yellow triangles. True corresponds to those travel times and horizontal derivatives evaluated by solving the eikonal equation numerically on a fine grid.

One of the most celebrated and well-known fact about NN is that it acts as a good universal function approximator (Hornik et al., 1989). Thus, it naturally provides us with multidimensional (determined by the number of input coordinates) nonlinear interpolation. Since the data NN learns the travel time for multiple sources and receivers characterized by their 3D location in space, the deep nature of the network function allows it to utilize most of the travel time information to predict travel times for sources and

receivers not represented in the training (interpolation). This feature is crucial to its effectiveness in the tomography application.

To exemplify our point, consider two measured data from two different event locations (Figures 2) in which the data are recorded along a certain mountain top. The data are synthetically generated using the Marmousi model (Figure 5a). The receiver arrays are distributed almost regularly, but with a large gap near the mountain top (around 3.5 km location). The sources (given by the blue dots) are sampled from the earthquake events provided by the catalog in (Fang et al., 2022), and are the ones used to train the data NN. The middle column in Figure 2 shows the predicted travel times, compared to the eikonal solution, at all points on the surface for a source given by the green star, top and bottom rows separately. We can see that the NN managed to capture the apex of the hyperbolic travel time curve quite accurately–even when no receivers are present in this area. More importantly, the lateral derivatives (Figure 2, c and f) predicted (by performing backpropagation on the NN) are also in good agreement with the one evaluated directly from the numerical solution of the eikonal solution using fine sampling.

3 Numerical Experiments

In this section, we examine the flexibility and stability of the proposed single-term PINNs' loss function (equation 12). The first subsection will cover the evaluation of the method in a 2D exploration seismological setting where the seismometers are placed along a borehole with varying data sparsity. To further demonstrate the flexibility of the formulation, we dedicate the second subsection for different recording surface settings in a surface tomography setup on a synthetic 2D velocity model. We will also demonstrate, in the third subsection, numerical reasoning behind the use of additive as opposed to multiplicative factorization for the travel time. Finally, we showcase the stability of the method in handling complex 3D velocity distribution. A single graphics processing unit NVIDIA Quadro RTX 8000 (48 GB memory) is used to perform the inversion (Data NN and PINNs' training).

3.1 Crosswell Tomography Experiment in 2D Medium

In this setting, we use a cropped 2D 1x1 km^2 SEG Advanced Modeling Program (SEAM) velocity model depicted in Figure 3a. We use a 0.01 km grid spacing in both (lateral and vertical) directions. The data are synthetically generated using a fast-marching-based eikonal solver and recorded at $x = 1$ km. We utilize an NN with 10 hidden layers having 10 neurons in each layer for the velocity NN and the data NN and a 20 hidden layers with 10 neurons in each layer for the travel time NN. All the three networks are initialized using the Glorot uniform initialization (Glorot & Bengio, 2010). The data NN is trained prior to the PINNs training using 500 epochs. The PINNs model is trained for 1000 epochs using the generated 112,211 training samples ($\mathbf{x} = \{x, z\}$) with a single (full) batch optimization. All the NNs use the Exponential Linear Unit (ELU) activation function for the hidden layers, the rectified linear unit (ReLU) is used on the last neuron of the velocity NN. The training of all data NN and the PINN is done with an Adam optimizer (Kingma & Ba, 2014). We use an initial learning rate of 0.00015 and decrease its value by half every 100 training epochs.

To highlight the ability of this framework in handling sparse data measurements, we share three scenarios differing in the receiver spacing. We consider three different receiver spacing of 0.02, 0.06, and 0.15 km. The sources are located (along the vertical z-direction) at x=0 km with a vertical spacing of 0.1 km starting at z=0 km. Figures 3 and 4 offer the cross-sectional view and vertical profiles for the different scenarios. From the inverted velocity section (Figure 3), we observe that PINNs are able to invert for the velocity accurately. Shown in Figure 3, we observe, as expected, a slight degradation with increasing receiver spacing in the inverted velocity's lateral variation. This degradation,

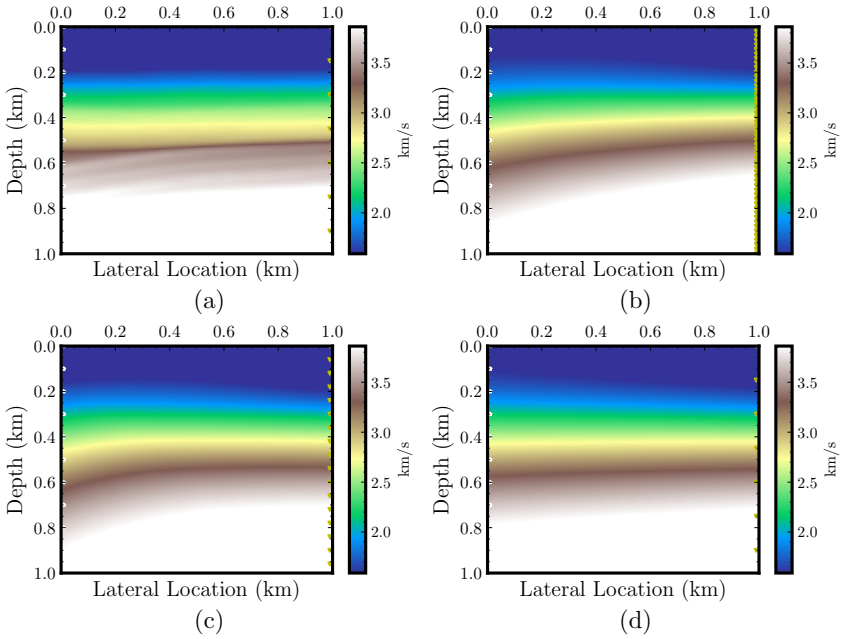

Figure 3. Inverted velocity sections for different receivers spacing (0.02, 0.06, and 0.15 km) in (b-d) comnpared to the the true velocity (a). The sources (white stars) and receivers (yellow triangles) are located at a fixed lateral location of 0 km and 1 km, respectively.

however, seems negligible as the inverted velocity on the largest receiver spacing still captures the shallow low-velocity layers of the model (Figure 4c) and follows the general increasing with depth velocity trend.

3.2 Surface Seismic Tomography in a 2D Medium

To further demonstrate the flexibility of the new travel time formulation (equation 5), we consider four distinct surface tomography tests with different data acquisition setups. These include the typical regularly sampled dense shot-receiver geometry, the same sampling scheme as in the first scenario with a gap, sparse shot distribution, and topography-dependent layout in a passive seismic recording setup. In all of these tests, we use an NN with 10 hidden layers containing 10 neurons in each layer for the velocity network and the data interpolation network. An NN with 20 hidden layers having 20 neurons in each layer is used for the higher dimensional travel time NN. We train the data NN for 5000 epochs. The travel time and velocity NNs use the ELU activation function while the data NN uses a rectified linear unit (ReLU) activation function. All of the three NNs use an Adam optimizer. We utilize a portion of the Marmousi model (Figure 5a) with a maximum lateral location of 8.6 km and maximum depth of 1 km. We discretize the computational domain using a 10 m vertical and 30 m lateral grid spacing.

Having the shot-receiver pairs regularly and densely sampled at a constant-depth acquisition setup, we consider the tomogram from the first scenario (Figure 5b) to be the reference solution for the other three scenarios. Specifically, the sources and receivers are sampled at a spacing interval of 200 m and 20 m, respectively. As shown in Figure 5b, the inversion manages to capture the general lateral variation accurately. We then use a sparser receiver sampling of 300 m and introduce a gap (in the sources and receivers, like an obstacle) that extends from 1.2 to 5.8 km lateral location in the model. Depicted in Figure 5c is the corresponding inverted velocity. We see degradation in terms of the

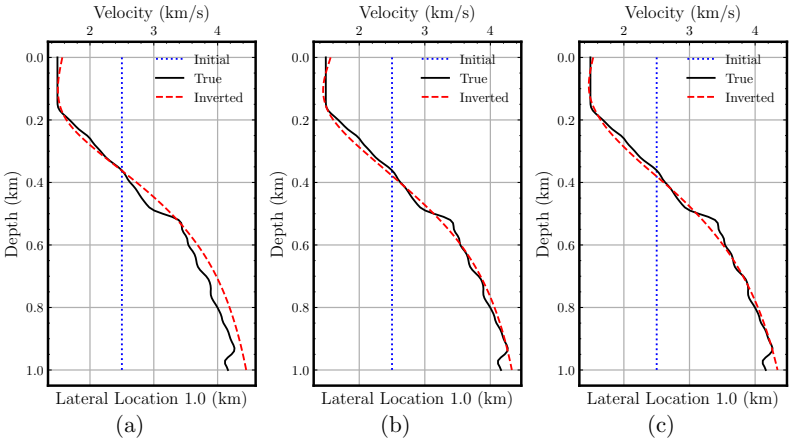

Figure 4. Vertical velocity profiles comparison for the cross-hole experiment for different receivers spacing (0.02, 0.06, and 0.15 km). Profiles in (a-c) are extracted at the borehole location (1 km lateral location) from Figure 3.

general lateral velocity resolution, especially under the gap, as we reduce almost half of the recording surface. The degraded tomogram, however, might still be considered a reasonable initial velocity model for further imaging (e.g., full waveform inversion). Using receivers sampled densely like in the first scenario, we reduce the shot sampling interval to 2.7 km in the third scenario. From the inverted velocity profile (Figure 5d), we can see that even with only four shots, the inverted velocity model captures the lateral variation of the actual model with high fidelity. Although some deep structures are obviously not well resolved, compared to the denser shot distribution (Figure 5b), the two tomograms are in good agreement.

To further highlight the efficacy of the new PINNs formulation, we test the same problem with a topography-dependent surface recording. It is straightforward to derive the PINNs' objective function by introducing a different ζ in equation 6, i.e., $\zeta(x, z)$. We also demonstrate the framework's ability to handle uneven source distributions by considering sampled earthquake locations from the southern part of California (Fang et al., 2022). The overall improvement from the tomogram can be attributed to the fact that now we image the transmission from the source as opposed to the diving waves in the previous three scenarios.

Finally, vertical velocity profiles are shown to further analyze the reconstructed velocity models. Although in general, the vertical profiles are identical (at least at the given lateral location), we can see that the 7-source experiments (Figures 5d and 6c) produce identical results when compared to the dense measurements (Figures 5b and 6a). The loss curves for the different scenarios are shown in Figure 7. We obtain slightly faster convergence with less data.

3.3 Suitable Factorization Approach for Data Hard-Constrained Eikonal Tomography

Before proceeding to the last sets of experiments, it is worth noting the role of travel time factorization on the PINNs' inversion. Specifically, the choice of factorization plays a major role in ensuring stability of the inversion. To this end, most of the previous works suggest the use of multiplicative factorization (i.e., changing the addition operation into

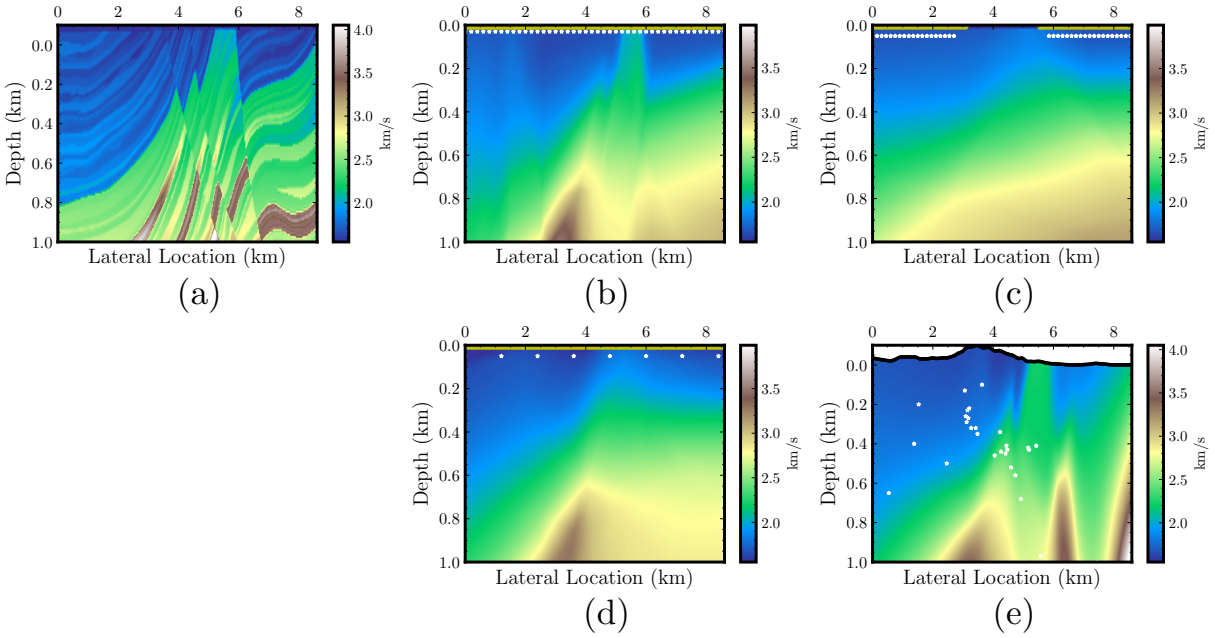

Figure 5. Inverted velocity sections for four different data acquisition scenarios (b-e) compared to the true velocity (a). The sources and receivers are denoted by the white and yellow dots, respectively.

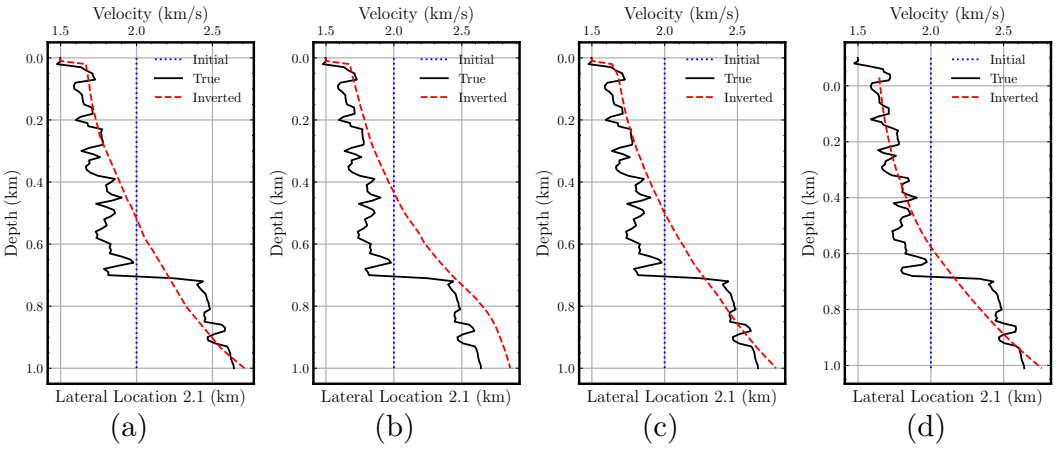

Figure 6. Vertical velocity profiles comparison for the 2D surface experiment for different scenarios in 5 (b-e). Profiles are extracted at 2.1 km lateral location.

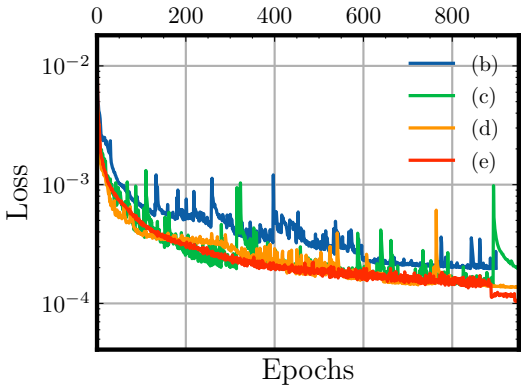

Figure 7. Loss curves for the training to obtain the inversion results shown in Figures 5(b-e).

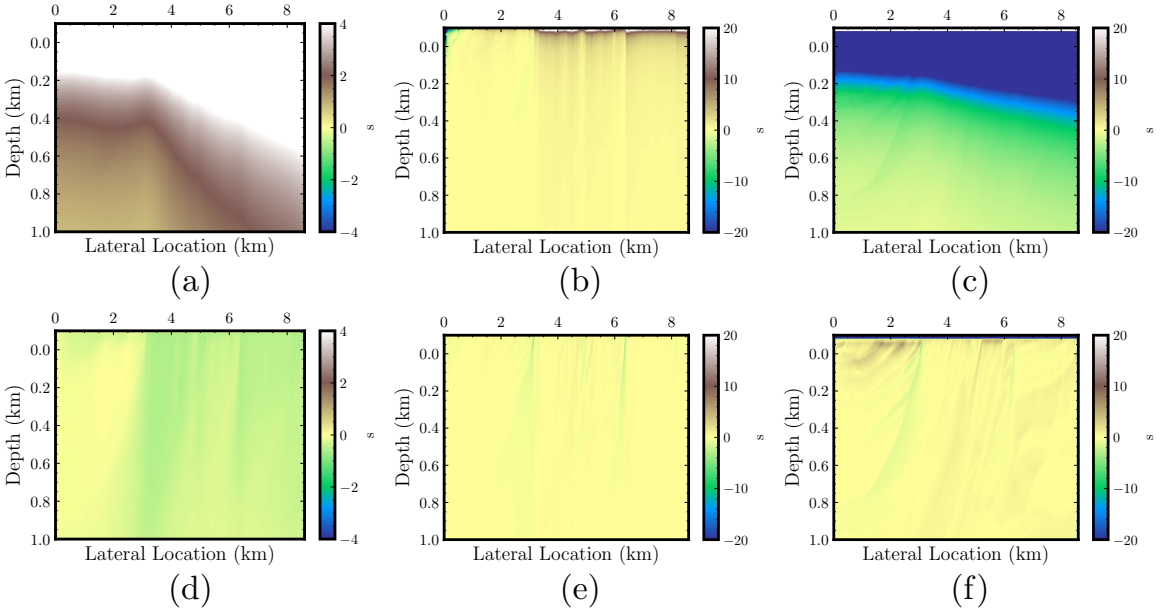

Figure 8. Factorization comparison. Multiplicative factorization, its lateral and vertical first-order derivatives extracted from the true $\hat{\tau}(\mathbf{x})$ are depicted in (a-c), respectively. The same configuration for the additive factorization approach are depicted in (d-f). The sections are extracted from the true $\hat{\tau}(\mathbf{x})$ field in Figure 5b.

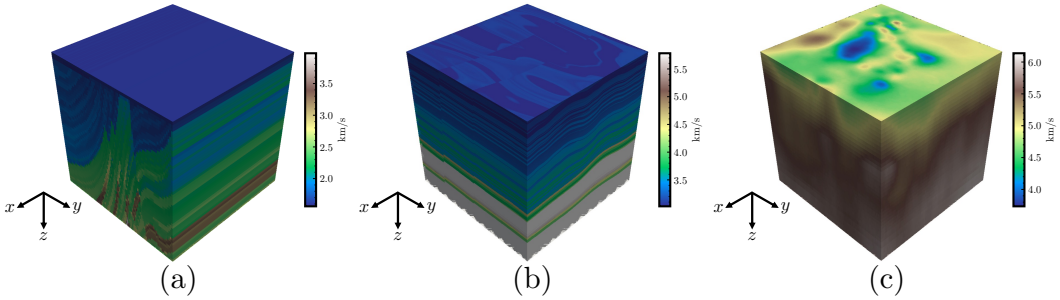

Figure 9. Three 3D velocity models used to simulate the travel time data. A 2.5D true velocity from a cropped Marmousi model (a). True 3D velocity models synthetic SEAM Phase II model (b) and derived from (White et al., 2021) (c).

multiplication in equation 2) to yield

$$T(\mathbf{x}) = \tau(\mathbf{x}) \cdot T_0(\mathbf{x}). \tag{13}$$

From our experiments, however, we find that for a hard-constraint PINN-based tomography, the additive factorization is more favorable than the multiplicative factorization. Figure 8, which represents a depth slice of the true $\hat{\tau}(\mathbf{x})$ for a surface tomography acquisition performed in the next section, exemplifies the point. These figures demonstrate that the additive factorization yields a much smoother true travel time factor ($\hat{\tau}(\mathbf{x})$) function. We can see that the $\hat{\tau}(\mathbf{x})$ values varies significantly between the depth near the source location ($z = 0.03$ km) and away from the source ($z = 1$ km) for the multiplicative factorization (Figure 8a), while the additive (Figure 8d) yields an almost constant function. The lateral and vertical (middle and right columns of Figure 8) further highlights the significant amplitude difference between the two approaches. Hence, with the known spectral bias feature of the NN (Rahaman et al., 2019), the travel time NN will favor predicting the additive (smoother) function over the multiplicative factorization function.

3.4 Surface Seismic Tomography in a 3D Medium

As mentioned earlier, the multi-term nature of previous PINN-based seismic tomography results in challenging training dynamics. Thus, previous PINN-based seismic tomography approaches, so far, only consider a 2D representation of the Earth. This section demonstrates the PINNs' ability to invert for a 3D velocity distribution by virtue of the more stable single-term loss function. We perform numerical simulation on three different 3D velocity models depicted in Figure 9.

3.4.1 2.5D Synthetic Velocity

We first consider a 2.5D velocity model. We duplicate the same cropped Marmousi model in the previous 2D subsection along the y axis depicted in Figure 9a. We use a 0.02 km grid spacing for the vertical and 0.1 km grid spacing for the lateral directions. The data are synthetically generated using a forward eikonal solver and recorded at $z = 0$ km. The sources are located near the surface and are regularly sampled with a lateral spacing of 1 km. We utilize a 12-layer with 12 neurons NN for the velocity NN and a 12-layer with 24 neurons for the travel time NN. All of these networks are initialized using the Glorot uniform initialization. These NNs use the ELU activation function and an Adam optimizer. The PINNs are trained for 1000 epochs using the generated 31,267,539 training samples ($\mathbf{x} = \{x, y, z\}$) with a 15,633 batch size. Throughout the subsequent experiments, we use an initial learning rate of 0.0005 and decrease its value by half ev-

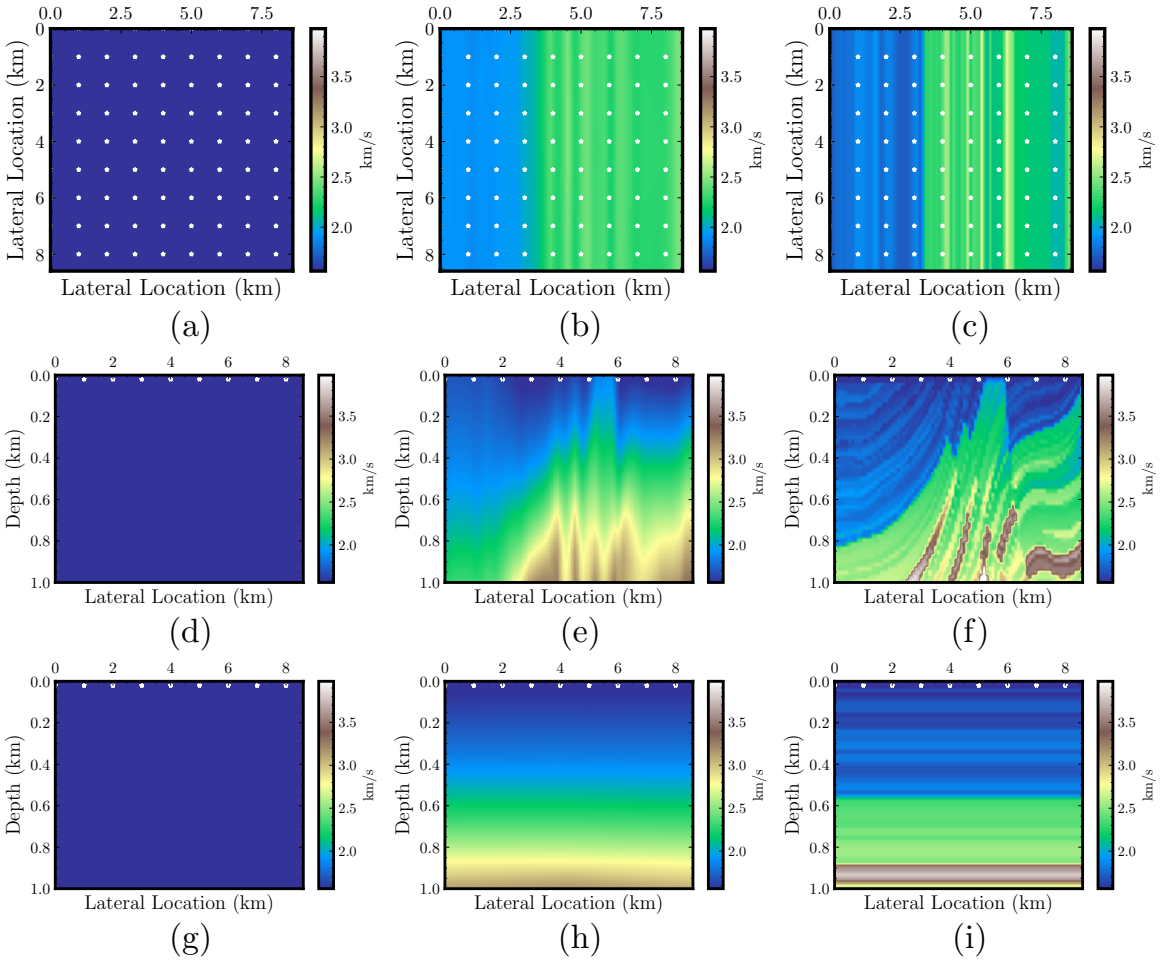

Figure 10. Velocity sections for the 2.5D cropped Marmousi model. The initial, inverted, and true velocity slices are depicted from the left to right columns. The top to bottom rows show the slices for xy, xz, yz planes on $z = 0.5$ km, $y = 3$ km, $x = 2$ km, respectively.

ery 200 training epochs. Further, we do not utilize the data NN for these regularly sampled measurements experiments.

Figure 10 offers a cross-sectional view of the inversion performance. We find a good agreement between the inverted velocity (middle column) and the true velocity model (right column). Figure 11 also demonstrates the accurate inverted velocity from a vertical profile view. Using the same computational domain and sample training points, we can further compare the 2D and 2.5D inversion process (Figure 12). By also comparing the inverted section between the 2D (Figure 5b) and 2.5D (Figure 10e), we observe a better and faster convergence of the 3D inversion. This is partially due to the fact that the rays cover a larger distance in the 2.5D model than the 2D case, and we have more illumination.

3.4.2 3D Synthetic Velocity

To investigate the framework's ability on a more realistic Earth model, we consider a synthetic experiment using the cropped 3D SEAM Phase II Arid model depicted in Figure 9b. We use a 0.0187 km grid spacing for the vertical and 0.075 km grid spacing

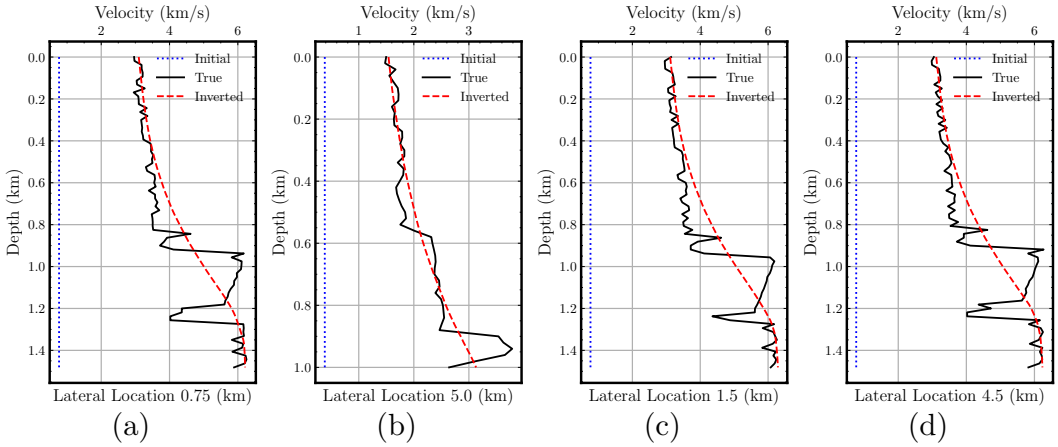

Figure 11. Vertical velocity profiles comparison for the 2.5D surface experiment at different lateral locations extracted from the inverted velocity in Figure 10.

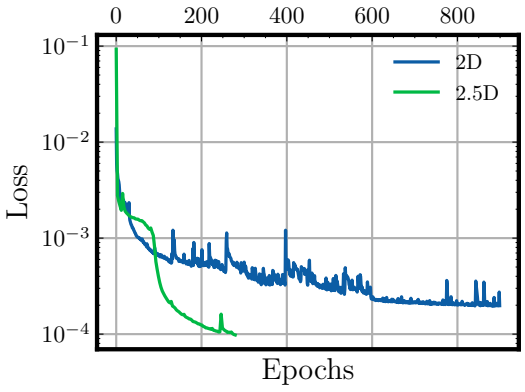

Figure 12. Loss curves comparison between the training for the inversion performed in Figure 5b and Figure 10e.

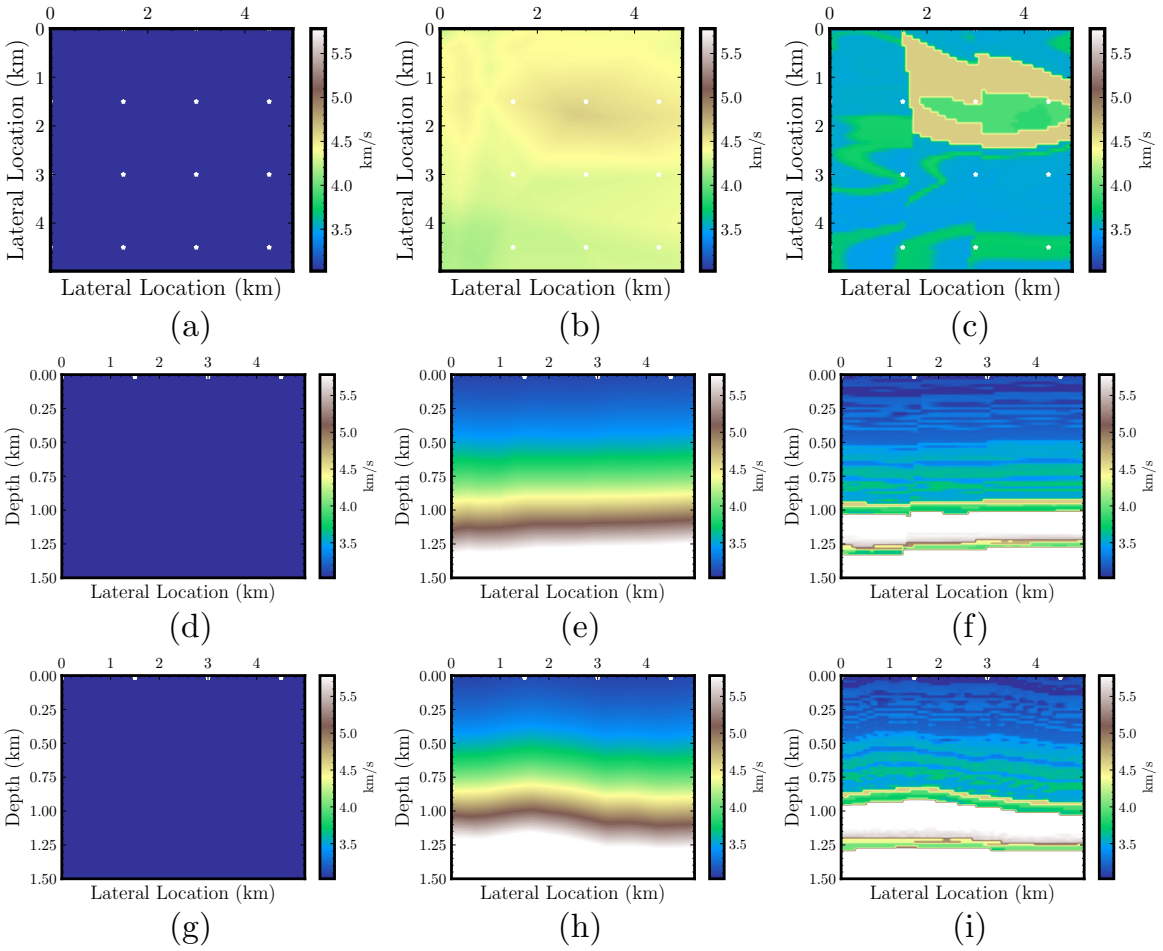

Figure 13. Velocity sections for the 3D SEAM Phase II Arid model. The initial, inverted, and true velocity slices are depicted from the left to right columns. The top to bottom row shows the slices for xy, xz, yz planes on $z = 0.84$ km, $y = 0.75$ km, $x = 2.25$ km, respectively.

for the lateral directions with a maximum lateral location and depth of 4.95 and 1.481 km, respectively. The data are synthetically generated using a forward eikonal solver and stored at $z = 0$ km. The sources are located near the surface and are regularly sampled with a lateral spacing of 1.5 km. The PINNs is trained for 1000 epochs using the generated 5,745,920 training samples ($\mathbf{x} = \{x, y, z\}$) with a 2,872 batch size. We utilize the same model complexity and training mechanism as in the previous 2.5D experiment.

The synthetic SEAM Phase II Arid is intended to mimic a challenging shallow crust (around the first 500 m) condition–typical to land exploration seismology cases. The right column of Figure 13 demonstrates the complex geological structures and lateral velocity variations of the model. The inverted velocity sections (middle column of Figure 13) show good agreement, in terms of the geological structures, with the input model. Both of the faulting systems (around 2 km lateral location in Figures 13e & f) as well as the anticlinal structure (around 2 km lateral location in Figures 13h & i) are recovered by the inversion process. The vertical profiles (Figure 14 also exemplifies the good inversion results as the inverted velocity captures the relatively constant velocity around the first 0.8 km depth and increases sharply at around 0.9 km depth.

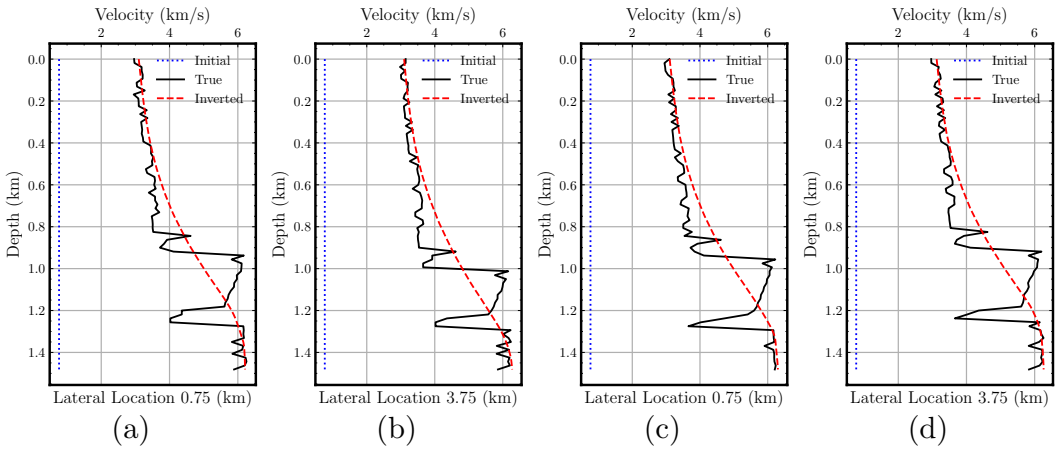

Figure 14. Vertical velocity profiles comparison for the 3D surface experiment at different lateral locations extracted from the inverted velocity in Figure 13.

3.4.3 3D Regional Velocity

Finally, to assess the performance of the framework's in dealing with a typical regional-scale eikonal tomography problem, we also perform a numerical experiment using inverted velocity provided from the study of (White et al., 2021) depicted in Figure 9c. It is worth mentioning that the purpose of this subsection is solely a further proof-of-concept that PINNs can as well be a viable option to deal with a typical regional-scale eikonal tomography. Compared to the previous 3D models, this velocity model contains more complex lateral velocity variation near the shallow crust. The data are synthetically generated using a forward eikonal solver and stored at $z = 0$ km. The sources are located near the surface and are regularly sampled with a lateral spacing of 32.637 km. The PINNs are trained for 1000 epochs using the generated 23,040,000 training samples ($\mathbf{x} = \{x, y, z\}$) with a 11,520 batch size. We utilize the same model complexity and training mechanism as in the previous 2.5D experiment.

Figure 16 illustrates the good agreement between the inverted and true (reference) velocity models. By comparing the inverted sections (middle column in Figure 16) and the true velocity (right column in Figure 16), we can see that the PINNs capture the complex shallow crust velocity variation. It is also worth noting that the inversion starts with a randomly initialized velocity model. This further demonstrates the stability of PINNs for eikonal tomography. Furthermore, we also find, at inference time (after the PINNs training), that we can instantly access the travel time field at a speed faster than a conventional eikonal solver (White et al., 2020). Running on the same Intel(R) Xeon(R) Gold 6230R CPU @ 2.10GHz, the trained PINNs model takes 3.68 minutes while the conventional eikonal solver takes 4.63 minutes.

4 Discussions

This work promotes a new formulation of a PINN-based eikonal tomography that relies on a single PDE residual term in its objective function. We reformulate the conventional eikonal tomography by replacing an eikonal solver with an NN that predicts the (factored) travel time field and another NN that substitutes the inversion kernel (e.g., adjoint-state method) to compute the velocity field. The minimization of the PINNs training will then be driven by obeying the PDE residual and a data misfit terms. In this case, the data measurement misfit is enforced using a hard constraint during the PINNs train-

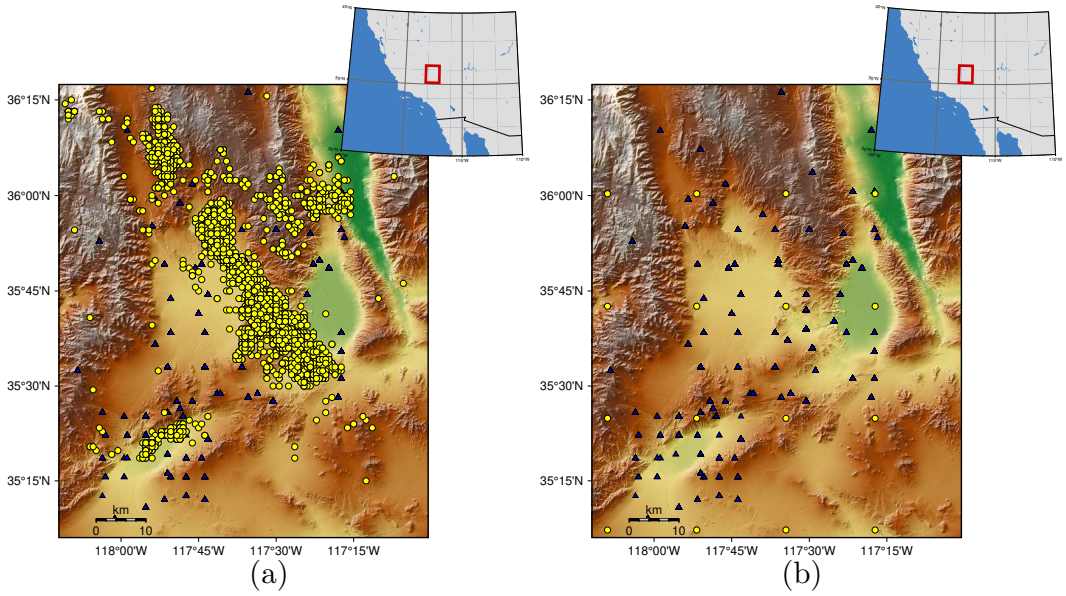

Figure 15. Map views of the events (yellow dots) and receivers (blue triangles) extracted from (White et al., 2021) (a) and the modified surface (z=0 km), regularly sampled, sources acquisition setup (b).

ing. Since lateral derivatives of the recorded travel time is needed, and to handle irregularly sampled measurements, we proposed the use of an NN to learn the measurements and inherit *tomographic* information from the source-receiver input location. Combining these three NNs as a framework, we found that hard-constrained measurement data significantly reduces the instability issues induced by the multi-term nature of the previous PINN-based eikonal tomography. These properties can be attributed to the natural inclusion of boundary conditions in the optimization problem. Thus, we alleviate a known pathology of the PINNs training in which a proper weighting scheme is needed to accommodate different loss function terms.

As a result, the single-term loss function improves the PINN-based eikonal tomography convergence. We found empirically that it can handle complex velocity distribution much better than the original PINN-based tomography formulation (Waheed et al., 2021a). It can even allow us to handle 3D measurements. Furthermore, the need for reliable measurement data can be partially achieved by a *tomographic* interpolation NN. We believe that by combining these components, PINN-based eikonal tomography can be a viable option for velocity model building at all scales. One benefit comes in the form that the trained velocity NN acts as a compressed representation of the velocity model (e.g., 3D velocity models). Another important feature is its flexibility in handling different acquisition scenarios. The method can still be applied without performing any coordinate transformation even for a topography-dependent recording surface. Moreover, as we have demonstrated, the travel time NN performs more efficiently than a conventional eikonal solver.

We have shown in the context of eikonal tomography that we can provide a stable inversion by reducing the number of loss terms in the training process. However, PINN-based approaches still require further development as an emerging option for further seismic tomography applications. Fortunately, the method presented here is orthogonal with previous attempts that view improper sampling as the major source of poor performance of PINNs. In other words, further development can accommodate a better sampling strat-

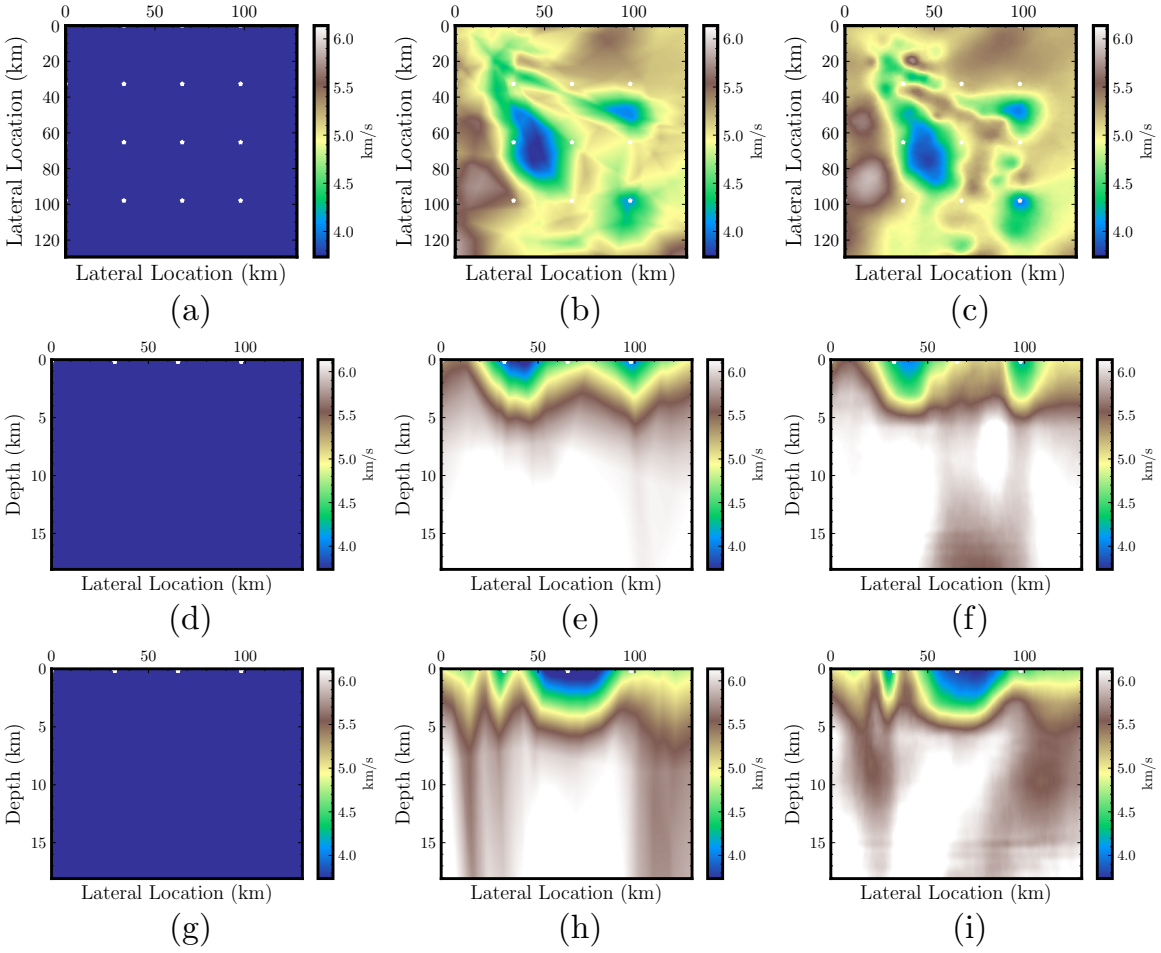

Figure 16. Velocity sections for the 3D Earth model from (White et al., 2021). The initial, inverted, and true velocity slices are depicted from the left to right columns. The top to bottom row shows the depth, latitude, and longitude sections on 0.91 km, 35.6° N, 117.8° W, respectively.

egy that, for example, obeys causality (Wang et al., 2022) and accounts for samples that have higher residual errors (Yu et al., 2022). We have also demonstrated the capability of PINNs to handle complex and realistic velocity models, the remaining challenge now is on how we further accelerate the training convergence, especially for large data measurements. Furthermore, for industrial-size 3D field data measurements, distributed multi-GPU training is inevitable. As in such a scenario, a single-GPU setup will run into the issue of out-of-memory problems. We anticipate that a multi-GPU setup with a mixed-even half-precision (to accelerate the convergence) can be made possible courtesy of the stable single-term loss function. Therefore, such a method can be a starting way forward to utilize PINNs for large-scale eikonal tomography.

5 Conclusions

We proposed a new framework for eikonal tomography, which is more suitable for a PINN-based tomographic workflow. We train, simultaneously from randomly initialized weights, two NNs to invert for the phase velocity and travel time field. We also suggest an optional data interpolation NN that is embedded with tomographic information. The new framework allows us to rely on a stable single-term loss function. The stability of the PINNs inversion can also be attributed to the additive factorization used to decompose the travel time field. We found that this factorization will result in a more stable PINN-based tomography compared to the multiplicative version. With rigorous numerical experimentation, we have demonstrated that it leads to a robust and flexible inversion framework that handles 3D Earth's model. Our new framework opens up new possibilities in dealing with sparse data measurements, that is typical in global seismological applications.

6 Data Availability Statement

Codes and data needed to reproduce the results presented here will be made available at https://github.com/hatsyim/HCPINNsEikonal.

References

Aki, K., Christoffersson, A., & Husebye, E. S. (1977). Determination of the three-dimensional seismic structure of the lithosphere. *Journal of Geophysical Research*, *82*(2), 277–296.

Alkhalifah, T. (2002, 7). Traveltime computation with the linearized eikonal equation for anisotropic media. *Geophysical Prospecting*, *50*, 373-382. Retrieved from `https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-2478.2002.00322.xhttps://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2478.2002.00322.xhttps://onlinelibrary.wiley.com/doi/10.1046/j.1365-2478.2002.00322.x` doi: 10.1046/J.1365-2478.2002.00322.X

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. (2018). Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research*, *18*(153), 1-43. Retrieved from `http://jmlr.org/papers/v18/17-468.html`

Bergman, B., Tryggvason, A., & Juhlin, C. (2004). High-resolution seismic traveltime tomography incorporating static corrections applied to a till-covered bedrock environment. *Geophysics*, *69*(4), 1082–1090.

Bording, R. P., Gersztenkorn, A., Lines, L. R., Scales, J. A., & Treitel, S. (1987). Applications of seismic travel-time tomography. *Geophysical Journal International*, *90*(2), 285–303.

Červený, V. (2000). *Seismic ray method*. Cambridge University Press.

Chen, Y., de Ridder, S. A., Rost, S., Guo, Z., Wu, X., & Chen, Y. (2022). Eikonal tomography with physics-informed neural networks: Rayleigh wave phase ve-

locity in the northeastern margin of the tibetan plateau. *Geophysical Research Letters*, *49*(21), e2022GL099053.

Dessa, J.-X., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., Uhira, K., & Kaneda, Y. (2004). Deep seismic imaging of the eastern nankai trough, japan, from multifold ocean bottom seismometer data by combined travel time tomography and prestack depth migration. *Journal of Geophysical Research: Solid Earth*, *109*(B2).

Dziewonski, A. M., Hager, B. H., & O'Connell, R. J. (1977). Large-scale heterogeneities in the lower mantle. *Journal of Geophysical Research*, *82*(2), 239–255.

Fang, H., White, M. C., Lu, Y., & Ben-Zion, Y. (2022). Seismic traveltime tomography of southern california using poisson-voronoi cells and 20 years of data. *Journal of Geophysical Research: Solid Earth*, *127*(5), e2021JB023307.

Fang, H., Zhang, H., Yao, H., Allam, A., Zigone, D., Ben-Zion, Y., ... van der Hilst, R. D. (2016). A new algorithm for three-dimensional joint inversion of body wave and surface wave data and its application to the southern california plate boundary region. *Journal of Geophysical Research: Solid Earth*, *121*(5), 3557–3569.

Fichtner, A. (2010). *Full seismic waveform modelling and inversion*. Springer Science & Business Media.

Fomel, S., Luo, S., & Zhao, H. (2009, 9). Fast sweeping method for the factored eikonal equation. *Journal of Computational Physics*, *228*, 6440-6455. doi: 10.1016/j.jcp.2009.05.029

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256).

Hauksson, E., Yang, W., & Shearer, P. M. (2012). Waveform relocated earthquake catalog for southern california (1981 to june 2011). *Bulletin of the Seismological Society of America*, *102*(5), 2239–2244.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*, 359-366.

Julian, B., Gubbins, D., et al. (1977). Three-dimensional seismic ray tracing. *Journal of Geophysics*, *43*(1), 95–113.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Klein, F. (2002). User's guide to hypoinverse-2000, a fortran program to solve for earthquake locations and magnitudes. open file report 02-171. *US Geological Survey*, 1–123.

Leung, S., & Qian, J. (2006). An adjoint state method for three-dimensional transmission traveltime tomography using first-arrivals.

Lin, F.-C., & Ritzwoller, M. H. (2011). Helmholtz surface wave tomography for isotropic and azimuthally anisotropic structure. *Geophysical Journal International*, *186*(3), 1104–1120.

Lin, F.-C., Ritzwoller, M. H., & Snieder, R. (2009). Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array. *Geophysical Journal International*, *177*(3), 1091–1110.

Lin, F.-C., Ritzwoller, M. H., Townend, J., Bannister, S., & Savage, M. K. (2007). Ambient noise rayleigh wave tomography of new zealand. *Geophysical Journal International*, *170*(2), 649–666.

Marsden, D. (1993). Static corrections—a review, part 1. *The leading edge*, *12*(1), 43–49.

McClenny, L., & Braga-Neto, U. (2020). Self-adaptive physics-informed neural networks using a soft attention mechanism. *arXiv preprint arXiv:2009.04544*.

Obrebski, M., Allen, R. M., Zhang, F., Pan, J., Wu, Q., & Hung, S.-H. (2012). Shear wave tomography of china using joint inversion of body and surface wave

constraints. *Journal of geophysical research: solid earth*, *117*(B1).

Plessix, R.-E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, *167*(2), 495–503.

Qin, F., Luo, Y., Olsen, K. B., Cai, W., & Schuster, G. T. (1992). Finite-difference solution of the eikonal equation along expanding wavefronts. *Geophysics*, *57*(3), 478–487.

Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., ... Courville, A. (2019). On the spectral bias of neural networks. In *International conference on machine learning* (pp. 5301–5310).

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019, 2). 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. doi: 10.1016/j.jcp.2018.10.045

Rawlinson, N., & Fishwick, S. (2012). Seismic structure of the southeast australian lithosphere from surface and body wave tomography. *Tectonophysics*, *572*, 111–122.

Rawlinson, N., Hauser, J., & Sambridge, M. (2008, 1). Seismic ray tracing and wavefront tracking in laterally heterogeneous media. *Advances in Geophysics*, *49*, 203-273. doi: 10.1016/S0065-2687(07)49003-3

Rawlinson, N., & Sambridge, M. (2004). Wave front evolution in strongly heterogeneous layered media using the fast marching method. *Geophysical Journal International*, *156*(3), 631–647.

Schiassi, E., Leake, C., Florio, M. D., Johnston, H., Furfaro, R., & Mortari, D. (2020, 5). Extreme theory of functional connections: A physics-informed neural network method for solving parametric differential equations. *arXiv preprint arXiv:2005.10632*. Retrieved from `https://arxiv.org/abs/2005.10632v1` doi: 10.48550/arxiv.2005.10632

Sethian, J. A. (1996). A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, *93*(4), 1591–1595. Retrieved from `https://www.pnas.org/content/93/4/1591` doi: 10.1073/pnas.93.4.1591

Simmons, N. A., Myers, S. C., Johannesson, G., & Matzel, E. (2012). Llnl-g3dv3: Global p wave tomography model for improved regional and teleseismic travel time prediction. *Journal of Geophysical Research: Solid Earth*, *117*(B10).

Simmons, N. A., Myers, S. C., Morency, C., Chiang, A., & Knapp, D. R. (2021). Spiral: a multiresolution global tomography model of seismic wave speeds and radial anisotropy variations in the crust and mantle. *Geophysical Journal International*, *227*(2), 1366–1391.

Smith, J. D., Azizzadenesheli, K., & Ross, Z. E. (2021). Eikonet: Solving the eikonal equation with deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(12), 10685-10696. doi: 10.1109/TGRS.2020.3039165

Taillandier, C., Noble, M., Chauris, H., & Calandra, H. (2009). First-arrival traveltime tomography based on the adjoint-state method. *Geophysics*, *74*. Retrieved from `http://library.seg.org/page/policies/terms` doi: 10.1190/1.3250266

Taufik, M. H., Waheed, U., & Alkhalifah, T. (2022). Upwind, no more: Flexible traveltime solutions using physics-informed neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-12. doi: 10.1109/TGRS.2022.3218754

Tavakoli, F. B., Operto, S., Ribodetti, A., & Virieux, J. (2017). Slope tomography based on eikonal solvers and the adjoint-state method. *Geophysical Journal International*, *209*(3), 1629–1647.

Thurber, C. H. (1983). Earthquake locations and three-dimensional crustal structure in the coyote lake area, central california. *Journal of Geophysical Research:*

Solid Earth, *88*(B10), 8226–8236.

Um, J., & Thurber, C. (1987). A fast algorithm for two-point seismic ray tracing. *Bulletin of the Seismological Society of America*, *77*(3), 972–986.

Vidale, J. (1988, 12). Finite-difference calculation of travel times. *Bulletin of the Seismological Society of America*, *78*(6), 2062-2076. Retrieved from `https://doi.org/10.1785/BSSA0780062062` doi: 10.1785/BSSA0780062062

Virieux, J., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, *74*(6), WCC1–WCC26.

Waheed, U., Alkhalifah, T., Haghighat, E., Song, C., & Virieux, J. (2021a, 4). Pintomo: Seismic tomography using physics-informed neural networks. *arXiv preprint arXiv:2104.01588*. doi: 10.48550/arxiv.2104.01588

Waheed, U., Haghighat, E., Alkhalifah, T., Song, C., & Hao, Q. (2021b, 10). Pineik: Eikonal solution using physics-informed neural networks. *Computers Geosciences*, *155*, 104833. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S009830042100131X` doi: 10.1016/j.cageo.2021.104833

Wang, S., Sankaran, S., & Perdikaris, P. (2022). Respecting causality is all you need for training physics-informed neural networks. *arXiv preprint arXiv:2203.07404*.

White, M. C., Fang, H., Catchings, R. D., Goldman, M. R., Steidl, J. H., & Ben-Zion, Y. (2021). Detailed traveltime tomography and seismic catalogue around the 2019 m w7. 1 ridgecrest, california, earthquake using dense rapid-response seismic data. *Geophysical Journal International*, *227*(1), 204–227.

White, M. C., Fang, H., Nakata, N., & Ben-Zion, Y. (2020). Pykonal: a python package for solving the eikonal equation in spherical and cartesian coordinates using the fast marching method. *Seismological Research Letters*, *91*(4), 2378–2389.

Yu, J., Lu, L., Meng, X., & Karniadakis, G. E. (2022). Gradient-enhanced physics-informed neural networks for forward and inverse pde problems. *Computer Methods in Applied Mechanics and Engineering*, *393*, 114823.

Zelt, C. A., Azaria, A., & Levander, A. (2006). 3d seismic refraction traveltime tomography at a groundwater contamination site. *Geophysics*, *71*(5), H67–H78.

Zelt, C. A., & Barton, P. J. (1998). Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the faeroe basin. *Journal of Geophysical Research: Solid Earth*, *103*(B4), 7187–7210.

Zhang, H., & Thurber, C. H. (2003). Double-difference tomography: The method and its application to the hayward fault, california. *Bulletin of the Seismological Society of America*, *93*(5), 1875–1889.

Zhang, J., & Toksöz, M. N. (1998). Nonlinear refraction traveltime tomography. *Geophysics*, *63*(5), 1726–1737.

Zhao, H. (2005). A fast sweeping method for eikonal equations. *Math. Comput.*, *74*, 603-627.