

Theory based machine learning TTI full waveform inversion based on recurrent neural network with the estimating of title angle

Tianze Zhang, Jian Sun, Kristopher Innanen, Daniel Trad

ABSTRACT

In this study, we use the recurrent neural network (RNN) to achieve TTI elastic full waveform inversion. The motivation for building such a network is that in real media full waveform inversion, the physics of wave propagation is very complex, and implementing insufficient accurate wave equations in such complex media would lead modeling errors. Most fractures are not vertically but with certain dips and azimuths, thus estimating the title angles along with the elastic parameters are important for accurately invert the parameters. The recurrent neural network (RNN) is a typical type of neural network that is consisted of several RNN cells. In this study, each RNN cell is designed according to the staggered grid stress velocity TTI wave equation and the Voigt stiffness parameters and the title angles are considered as the parameters in this inversion. Based on the forward computational graph, the gradients with respect to each parameter are given by the backpropagation of the forward computational graph. In order to mitigate the cross talk, we use high order total variation (TV) regularization to mitigate the cross-talk in the inversion, Numerical inversions using simple models and complex models prove the validation of this method.

INTRODUCTION

Full-waveform inversion (FWI), an iterative nonlinear optimization procedure, can be used to obtain high-resolution seismic images. Tarantola (1984) shows that the Fréchet derivatives of a waveform-difference misfit function can be computed via the interaction between a forward-propagating wavefield and a reverse-propagating data-residual wavefield. In TI media the anisotropic symmetry axis is commonly chosen to be orthogonal to the bedding plane of geologic structures. Wang and Tsvankin (2013) shows that the anisotropic parameters α_0 , ϵ , and δ are recoverable through the moveout of a common-image-gather of Kirchhoff migration, which is iterated using ray-based tomography. The symmetry axis is assumed to be perpendicular to reflectors in the migrated image. Oropenza and McMechan (2014) estimate the 2D acoustic TTI parameters α_0 , ϵ , and δ by maximizing the stacked amplitude along the traveltimes curve in a common-reflection-point gather. During the inversion, α_0 is weighted less than θ and ϵ to reduce the dominance of α_0 . The orientation of the symmetry axis orthogonal to the local reflector orientation is calculated using parsimonious migration Hua and McMechan (2003), and errors in θ are shown to affect the recovery of ϵ and δ . In 2D elastic VTI media, Lee et al. (2010) propose a strategy, in which the Voigt parameters C_{11} , C_{33} , C_{44} , and ϵ are independently updated, and C_{11} is updated using the estimated C_{33} and ϵ . Gholami et al. (2013) consider three different parameterizations, namely, α_0 , δ , ϵ , and α_{NMO} , δ , ϵ , and α_{NMO} , δ , n . In each of the above parameterizations, the dimensionless parameters are less influential than the wave speed α_0 or α_{NMO} ; at the final iteration, the inverted Thomsen parameters are similar to the initial ones. These experiments imply that keeping the true Thomsen parameters for example, obtained from well logs, Hornby et al. (2003), fixed during the inversion does not signifi-

cantly affect the recovery of the wave speeds. Gholami et al. (2013) also note that C_{11} and C_{33} significantly affect the data, and they are three times more sensitive than C_{14} , which is subject to trade-offs for far offset data.

The power of machine learning has become more and more noticed by Geophysicists. One of the most powerful advantages for machine learning is that it could build the linear or nonlinear relationship between the input and output through training the trainable parameters. Data-driven methods in machine learning has been implemented by many Geophysics researchers. Reading et al. (2015) showed how the statistical, robust, output from the machine learning exercise can be used to guide the construction of improved volume geometry within a 3D GOCAD geological and geophysical modeling environment. Cracknell and Reading (2013) found that the use of optimal uncertainty thresholds significantly improves the overall classification accuracy of RF predictions, but not those of supportive vector machine (SVM), by eliminating the maximum number of incorrectly classified samples while preserving the maximum number of correctly classified samples Wrona et al. (2018) used state-of-the-art 3D broadband seismic reflection data of the northern North Sea. Yang and Ma (2019) introduced an inversion method that is based on the convolutional neural network, which reduces the strong dependency of initial models for scalar wave FWI. Lin and Wu (2018) designed the InverseNet bases on the convolutional network to do full waveform inversion. However, these methods are mainly based on the data-driven methods, which means that these methods need a huge amount of time for training and ignore the theory we know about inversion and wave propagation. If the over-fitting and under-fitting problems occur in these methods, the inversion results would be badly influenced.

In this study, we are implementing theory-guided networks, which are different from the data-driven methods. Data-driven methods use a large amount of data set to train a network so that a network can learn the pattern between the input and output and build a mapping between them. However, such a training process ignores the basic physics laws we know about wave propagation. Also, how to choose the training data set would be an issue since the wave propagation in the under-ground world is so complex that we can hardly get the exact mapping between the seismic records and the elastic parameters in the real data set for training. It has recently been shown (Sun et al., 2020) that seismic wave propagation can be simulated with a specialized recurrent neural network (RNN), and that the process of training such a network with a single seismic data set is equivalent to carrying out seismic full waveform inversion (FWI). The recurrent neural network in this test is built by using such a network the RNN cell is designed according to the staggered grid TTI wave equation.

This article is organized as follows. First, we introduce the how TTI elastic wave equation is derived. Second, we show how the title angle can influence the waveform. Third, we introduce the high order total variation misfits. Fourth, we perform inversions on simple and complex models to prove the validation of this method. The final part is the conclusions.

DERIVATION OF THE TTI WAVE EQUATION

Tilted transverse isotropy (TTI) is a form of rotated transverse isotropy that can be described by VTI parameters together with two spatially variable rotation angles, namely, azimuth ϕ and tilt angle θ . For 2D TTI, the orientation of the anisotropic symmetry axis can be described using only the tilt angle. In real earth medium, the patterns of the developed fractures are very complex. Most fractures are not vertically but with certain dips and azimuths. In this case, the fractured medium is equivalent to an elastic TTI medium and its stiffness matrix in Cartesian coordinate is no longer the same as the VTI medium constitutive equations in the constitutive coordinate. For TTI medium with a tilted axis of symmetry, there is an angle between its constitutive coordinates and the observation coordinates.

Using bond transformation, the elasticity coefficients of TTI medium with any azimuths and dips can be transformed from the constitutive coordinates to the Cartesian coordinates. Subsequent wave equation derivation and solution can be facilitated. Bond transformation matrix includes the polarization transformation matrix \mathbf{M}_θ and the orientation transformation matrix \mathbf{M}_ϕ :

$$\mathbf{M}_\theta = \begin{bmatrix} \cos^2\theta & 0 & \sin^2\theta & 0 & -\sin 2\theta & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \sin^2\theta & 0 & \cos^2\theta & 0 & \sin 2\theta & 0 \\ 0 & 0 & 0 & \cos\theta & 0 & \sin\theta \\ \frac{1}{2}\sin 2\theta & 0 & -\frac{1}{2}\sin 2\theta & 0 & \cos 2\theta & 0 \\ 0 & 0 & 0 & -\sin\theta & 0 & \cos\theta \end{bmatrix} \quad (1)$$

$$\mathbf{M}_\phi = \begin{bmatrix} \cos^2\phi & \sin^2\phi & 0 & 0 & 0 & -\sin 2\phi \\ \sin^2\phi & \cos^2\phi & 0 & 0 & 0 & \sin 2\phi \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cos\phi & \sin\phi & 0 \\ 0 & 0 & 0 & -\cos\phi & \cos\phi & 0 \\ -\frac{1}{2}\sin 2\phi & -\frac{1}{2}\sin 2\phi & 0 & 0 & 0 & \cos 2\phi \end{bmatrix} \quad (2)$$

In Cartesian observation system, the elastic coefficient matrix of a TTI media can be expressed as:

$$\mathbf{C}_{\text{TTI}} = \mathbf{M}_\theta \mathbf{M}_\theta \mathbf{C}_{\text{VTI}} \mathbf{M}_\phi^T \mathbf{M}_\phi^T, \quad (3)$$

where the superscript T denotes the matrix transpose. In this study we note rotation matrix \mathbf{R} as rotation matrix and $\mathbf{R} = \mathbf{M}_\theta \mathbf{M}_\phi$.

The transformation matrix is:

$$\mathbf{M}_\theta \mathbf{M}_\theta = \begin{bmatrix} \cos^2\phi & \sin^2\phi \cos^2\theta & \sin^2\phi \sin^2\theta & -\sin^2\phi \sin 2\theta & \sin 2\phi \sin\theta & -\sin 2\phi \cos\theta \\ \sin^2\phi & \cos^2\phi \cos^2\theta & \cos^2\phi \sin^2\theta & -\cos^2\phi \sin 2\theta & -\sin 2\phi \sin\theta & \sin 2\phi \cos\theta \\ 0 & \sin^2\theta & \cos^2\theta & \sin 2\theta & 0 & 0 \\ 0 & \frac{1}{2} \cos\phi \sin 2\theta & -\frac{1}{2} \cos\phi \sin 2\theta & \cos\phi \cos 2\theta & \sin\phi \cos\theta & \sin\phi \sin\theta \\ 0 & -\frac{1}{2} \sin\phi \sin 2\theta & \frac{1}{2} \sin\phi \sin 2\theta & -\sin\phi \cos 2\theta & \cos\phi \cos\theta & \cos\phi \sin\theta \\ \frac{1}{2} \sin 2\phi & -\frac{1}{2} \sin 2\phi \cos^2\theta & -\frac{1}{2} \sin 2\phi \sin^2\theta & \frac{1}{2} \sin 2\phi \sin 2\theta & -\cos 2\phi \sin\theta & \cos 2\phi \cos\theta \end{bmatrix}, \quad (4)$$

Because we are stating the 2D theory, only a tilt-angle rotation is involved. If a 2D elastic VTI medium is considered, the symmetric stiffness (Voigt) matrix is reduced to:

$$\mathbf{C}_{\text{VTI}} = \begin{bmatrix} C_{11} & C_{13} & 0 \\ C_{13} & C_{33} & 0 \\ 0 & 0 & C_{44} \end{bmatrix}. \quad (5)$$

The general 2D case the rotation matrix \mathbf{R} :

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{13} & R_{15} \\ R_{31} & R_{33} & R_{35} \\ R_{51} & R_{53} & R_{55} \end{bmatrix} = \begin{bmatrix} \cos^2\phi \cos^2\theta & \cos^2\phi \sin^2\theta & \cos^2\phi \sin 2\theta \\ \sin^2\theta & \cos^2\theta & \sin 2\theta \\ \frac{\cos\phi \sin 2\theta}{2} & -\frac{\cos\phi \sin 2\theta}{2} & \cos\phi \cos 2\theta \end{bmatrix}. \quad (6)$$

The stiffness matrix becomes:

$$\mathbf{R} \mathbf{C}_{\text{VTI}} \mathbf{R}^T = \begin{bmatrix} R_{11}C_{11} + R_{13}C_{13} & R_{11}C_{13} + R_{13}C_{33} & R_{15}C_{55} \\ R_{31}C_{11} + R_{33}C_{13} & R_{31}C_{13} + R_{33}C_{33} & R_{35}C_{55} \\ R_{51}C_{11} + R_{53}C_{13} & R_{51}C_{13} + R_{53}C_{33} & R_{55}C_{55} \end{bmatrix} \begin{bmatrix} R_{11} & R_{31} & R_{51} \\ R_{13} & R_{33} & R_{53} \\ R_{15} & R_{35} & R_{55} \end{bmatrix} = \begin{bmatrix} \hat{C}_{11} & \hat{C}_{13} & \hat{C}_{15} \\ \hat{C}_{13} & \hat{C}_{33} & \hat{C}_{35} \\ \hat{C}_{15} & \hat{C}_{35} & \hat{C}_{55} \end{bmatrix}, \quad (7)$$

where the \hat{C}_{11} , \hat{C}_{13} , \hat{C}_{15} , \hat{C}_{33} , \hat{C}_{35} , \hat{C}_{55} are the stiffness coefficients for the TTI media. The expression for each stiffness matrix parameter can be expressed as:

$$\begin{bmatrix} R_{11}(R_{11}C_{11} + R_{13}C_{13}) + R_{13}(R_{11}C_{13} + R_{13}C_{33}) + R_{15}C_{55}R_{15} \\ (R_{11}C_{11} + R_{13}C_{13})R_{31} + (R_{11}C_{13} + R_{13}C_{33})R_{33} + R_{15}C_{55}R_{35} \\ (R_{11}C_{11} + R_{13}C_{13})R_{51} + (R_{11}C_{13} + R_{13}C_{33})R_{53} + R_{15}C_{55}R_{55} \\ (R_{31}C_{11} + R_{33}C_{13})R_{11} + (R_{31}C_{13} + R_{33}C_{33})R_{13} + R_{35}C_{55}R_{15} \\ (R_{31}C_{11} + R_{33}C_{13})R_{31} + (R_{31}C_{13} + R_{33}C_{33})R_{33} + R_{35}C_{55}R_{35} \\ (R_{31}C_{11} + R_{33}C_{13})R_{51} + (R_{31}C_{13} + R_{33}C_{33})R_{53} + R_{35}C_{55}R_{55} \\ (R_{51}C_{11} + R_{53}C_{13})R_{11} + (R_{51}C_{13} + R_{53}C_{33})R_{13} + R_{55}C_{55}R_{15} \\ (R_{51}C_{11} + R_{53}C_{13})R_{31} + (R_{51}C_{13} + R_{53}C_{33})R_{33} + R_{55}C_{55}R_{35} \\ (R_{51}C_{11} + R_{53}C_{13})R_{51} + (R_{51}C_{13} + R_{53}C_{33})R_{53} + R_{55}C_{55}R_{55} \end{bmatrix} = \begin{bmatrix} \hat{C}_{11} \\ \hat{C}_{13} \\ \hat{C}_{15} \\ \hat{C}_{13} \\ \hat{C}_{33} \\ \hat{C}_{35} \\ \hat{C}_{15} \\ \hat{C}_{35} \\ \hat{C}_{55} \end{bmatrix}. \quad (8)$$

The constitutive equations in a viscoelastic TTI fractured medium are:

$$\begin{bmatrix} \sigma_{xx} \\ \sigma_{zz} \\ \sigma_{xz} \end{bmatrix} = \begin{bmatrix} \hat{C}_{11} & \hat{C}_{13} & \hat{C}_{15} \\ \hat{C}_{13} & \hat{C}_{33} & \hat{C}_{35} \\ \hat{C}_{15} & \hat{C}_{35} & \hat{C}_{55} \end{bmatrix} \begin{bmatrix} v_{xx} \\ v_{zz} \\ v_{xz} \end{bmatrix} \quad (9)$$

Detail expression of the normal stress, σ_{xx} and σ_{zz} , and shear stress σ_{xz} can be expressed as:

$$\begin{aligned}\sigma_{xx} = & [R_{11}(R_{11}C_{11} + R_{13}C_{13}) + R_{13}(R_{11}C_{13} + R_{13}C_{33}) + R_{15}C_{55}R_{15}]v_{xx} \\ & + [(R_{11}C_{11} + R_{13}C_{13})R_{31} + (R_{11}C_{13} + R_{13}C_{33})R_{33} + R_{15}C_{55}R_{35}]v_{zz} \\ & + [(R_{11}C_{11} + R_{13}C_{13})R_{51} + (R_{11}C_{13} + R_{13}C_{53})R_{53} + R_{15}C_{55}R_{55}]v_{xz}\end{aligned}\quad (10)$$

$$\begin{aligned}\sigma_{zz} = & [(R_{31}C_{11} + R_{13}C_{13})R_{11} + (R_{31}C_{13} + R_{33}C_{33})R_{13} + R_{35}C_{55}R_{15}]v_{xx} \\ & + [(R_{31}C_{11} + R_{33}C_{13})R_{31} + (R_{31}C_{13} + R_{33}C_{33})R_{33} + R_{35}C_{55}R_{35}]v_{zz} \\ & + [(R_{31}C_{11} + R_{33}C_{13})R_{51} + (R_{31}C_{13} + R_{33}C_{33})R_{53} + R_{35}C_{55}R_{55}]v_{xz}\end{aligned}\quad (11)$$

$$\begin{aligned}\sigma_{xz} = & [(R_{51}C_{11} + R_{53}C_{13})R_{11} + (R_{51}C_{13} + R_{53}C_{33})R_{13} + R_{55}C_{55}R_{15}]v_{xx} \\ & + [(R_{51}C_{11} + R_{53}C_{13})R_{31} + (R_{51}C_{13} + R_{53}C_{33})R_{33} + R_{55}C_{55}R_{35}]v_{zz} \\ & + [(R_{51}C_{11} + R_{53}C_{13})R_{51} + (R_{51}C_{13} + R_{53}C_{33})R_{53} + R_{55}C_{55}R_{55}]v_{xz}\end{aligned}\quad (12)$$

The geometric equations are expressed as:

$$\begin{aligned}v_{xx} &= \frac{\partial v_x}{\partial x}, \\ v_{zz} &= \frac{\partial v_z}{\partial z}, \\ v_{xz} &= \frac{\partial v_x}{\partial z},\end{aligned}\quad (13)$$

$$\begin{aligned}\frac{\partial v_x}{\partial t} &= \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xz}}{\partial z}, \\ \frac{\partial v_z}{\partial t} &= \frac{\partial \sigma_{zz}}{\partial z} + \frac{\partial \sigma_{xz}}{\partial x},\end{aligned}\quad (14)$$

In this study, we formulate the recurrent neural network (RNN) to simulate waveform inversion. Algorithm 1 shows the basic structure of the recurrent neural network. At each time step the discrete sources s_x and s_z act as inputs; the velocity and stress information, $v_x^{t-\frac{1}{2}}$, $v_z^{t-\frac{1}{2}}$, σ_{xx}^t , σ_{zz}^t , and σ_{xz}^t , is communicated between the RNN cells; the partial derivative fields, $\partial_x \sigma_{xx}^t$, $\partial_z \sigma_{zz}^t$, $\partial_x \sigma_{xz}^t$, $\partial_z \sigma_{xz}^t$, $\partial_x v_x^{t+\frac{1}{2}}$, $\partial_z v_x^{t+\frac{1}{2}}$, $\partial_x v_z^{t+\frac{1}{2}}$, $\partial_z v_z^{t+\frac{1}{2}}$ are the internal variables in each RNN cell.

Algorithm 1 Sequence of calculations in the RNN cell

Input: Source: s_x , s_z ; velocity and stress fields at the previous time step. TTI Voigt stiffness parameters: $\hat{C}_{11}, \hat{C}_{13}, \hat{C}_{15}, \hat{C}_{33}, \hat{C}_{35}, \hat{C}_{15}, \hat{C}_{35}, \hat{C}_{55}$,

Output: Update velocity field at $t + \frac{1}{2}$ and stress fields at $t + 1$

$$1: \sigma_{xx}^t \leftarrow \sigma_{xx}^t + s_x$$

- 2: $\sigma_{zz}^t \leftarrow \sigma_{zz}^t + s_z$
- 3: $\partial_x \sigma_{xx}^t \leftarrow (\sigma_{xx}^t * \mathbf{k}_{x_1}) / \rho$
- 4: $\partial_z \sigma_{xz}^t \leftarrow (\sigma_{xz}^t * \mathbf{k}_{z_2}) / \rho$
- 5: $\partial_x \sigma_{xz}^t \leftarrow (\sigma_{xz}^t * \mathbf{k}_{x_2}) / \rho$
- 6: $\partial_z \sigma_{zz}^t \leftarrow (\sigma_{zz}^t * \mathbf{k}_{z_1}) / \rho$
- 7: $v_x^{t+\frac{1}{2}} \leftarrow v_x^{t-\frac{1}{2}} + dt(\partial_x \sigma_{xx}^t) + dt(\partial_z \sigma_{xz}^t)$
- 8: $v_z^{t+\frac{1}{2}} \leftarrow v_z^{t-\frac{1}{2}} + dt(\partial_z \sigma_{zz}^t) + dt(\partial_x \sigma_{xz}^t)$
- 9: $\partial_x v_x^{t+\frac{1}{2}} \leftarrow v_x^{t+\frac{1}{2}} * \mathbf{k}_{x_2}$
- 10: $\partial_z v_x^{t+\frac{1}{2}} \leftarrow v_x^{t+\frac{1}{2}} * \mathbf{k}_{z_1}$
- 11: $\partial_x v_z^{t+\frac{1}{2}} \leftarrow v_z^{t+\frac{1}{2}} * \mathbf{k}_{x_1}$
- 12: $\partial_z v_z^{t+\frac{1}{2}} \leftarrow v_z^{t+\frac{1}{2}} * \mathbf{k}_{z_2}$
- 13: $\sigma_{xx}^t \leftarrow \sigma_{xx}^{t-1} + dt \left(\hat{C}_{11} \partial_x v_x^{t+\frac{1}{2}} + \hat{C}_{13} \partial_z v_z^{t+\frac{1}{2}} + \hat{C}_{15} \partial_z v_{xz}^{t+\frac{1}{2}} \right)$
- 14: $\sigma_{zz}^t \leftarrow \sigma_{zz}^{t-1} + dt \left(\hat{C}_{13} \partial_x v_x^{t+\frac{1}{2}} + \hat{C}_{33} \partial_z v_z^{t+\frac{1}{2}} + \hat{C}_{35} \partial_z v_{xz}^{t+\frac{1}{2}} \right)$
- 15: $\sigma_{xz}^t \leftarrow \sigma_{xz}^{t-1} + dt \left(\hat{C}_{15} \partial_x v_x^{t+\frac{1}{2}} + \hat{C}_{35} \partial_z v_z^{t+\frac{1}{2}} + \hat{C}_{55} \partial_z v_{xz}^{t+\frac{1}{2}} \right)$

Algorithm 2 Loop for TTI elastic RNN FWI

- 1: Set trainable parameters: $\mathbf{C}_{11}, \mathbf{C}_{13}, \mathbf{C}_{33}, \mathbf{C}_{44}, \theta$ in this test.
 - 2: Set optimizers for parameters: $Optimizer_1, Optimizer_2, Optimizer_3, Optimizer_4, Optimizer_5$ for $\mathbf{C}_{11}, \mathbf{C}_{13}, \mathbf{C}_{33}, \mathbf{C}_{44}, \theta$ respectively.
 - 3: **for** $iter \in [1, maxiter]$ or not converge **do**
 - 4: $D_{syn} = \text{RNN}(\mathbf{C}_{11}, \mathbf{C}_{13}, \mathbf{C}_{33}, \mathbf{C}_{44}, \theta)$: generate synthetic data
 - 5: $loss = \text{costFunc}(D_{syn}, D_{obs})$: calculate misfits
 - 6: $loss.backward()$: Backpropagation and give gradients for the parameters
 - 7: $optimizers.step()$: update parameters
 - 8: **end for**
-

The * symbol represents the machine learning image convolution operator. This image convolution is the process of adding each element of the image to its local neighbors, weighted by the image convolution kernel. We find that this image convolution operator is also capable of calculating space partial derivatives if the convolution kernel is designed according to the finite difference coefficients. dx, dz are the grid intervals, and the image convolution kernels are: $\mathbf{k}_{x_1} = \mathbf{a}/dx, \mathbf{k}_{x_2} = \mathbf{b}/dx, \mathbf{k}_{z_1} = \mathbf{a}^T/dz$, and $\mathbf{k}_{z_2} = \mathbf{b}^T/dz$, where $\mathbf{a} = [0, 1/24, -9/8, 9/8, -1/24]$ and $\mathbf{b} = [1/24, -9/8, 9/8, -1/24, 0]$. \mathbf{a} and \mathbf{b} are 1×5 dimension arrays. \mathbf{k}_{x_1} and \mathbf{k}_{x_2} are kernels, for the image convolution process, responsible for calculating the staggered grid space partial derivative in x direction. \mathbf{k}_{z_1} and \mathbf{k}_{z_2} are kernels, for the image convolution process, responsible for calculating the staggered grid space partial derivative in z direction, and that is also why the arrays, \mathbf{a} and \mathbf{b} , are transposed in \mathbf{k}_{z_1} and \mathbf{k}_{z_2} . Space partial derivative calculated in this way is, mathematically, the same with conventional staggered grid method. $\hat{C}_{11}, \hat{C}_{13}, \hat{C}_{15}, \hat{C}_{13}, \hat{C}_{33}, \hat{C}_{35}, \hat{C}_{15}, \hat{C}_{35}, \hat{C}_{55}$, are the stiffness (Voigt) matrix in TTI media.

Figure 1 shows the snapshots for a TTI model. In Figure 1 $C_{11} = 9$ GPa, $C_{13} = 1.79$ GPa, $C_{33} = 8$ GPa, $C_{44} = 2.79$ GPa. In Figure 1 (a)-(d) the title angle is $\theta = \frac{\pi}{2}$. Figure

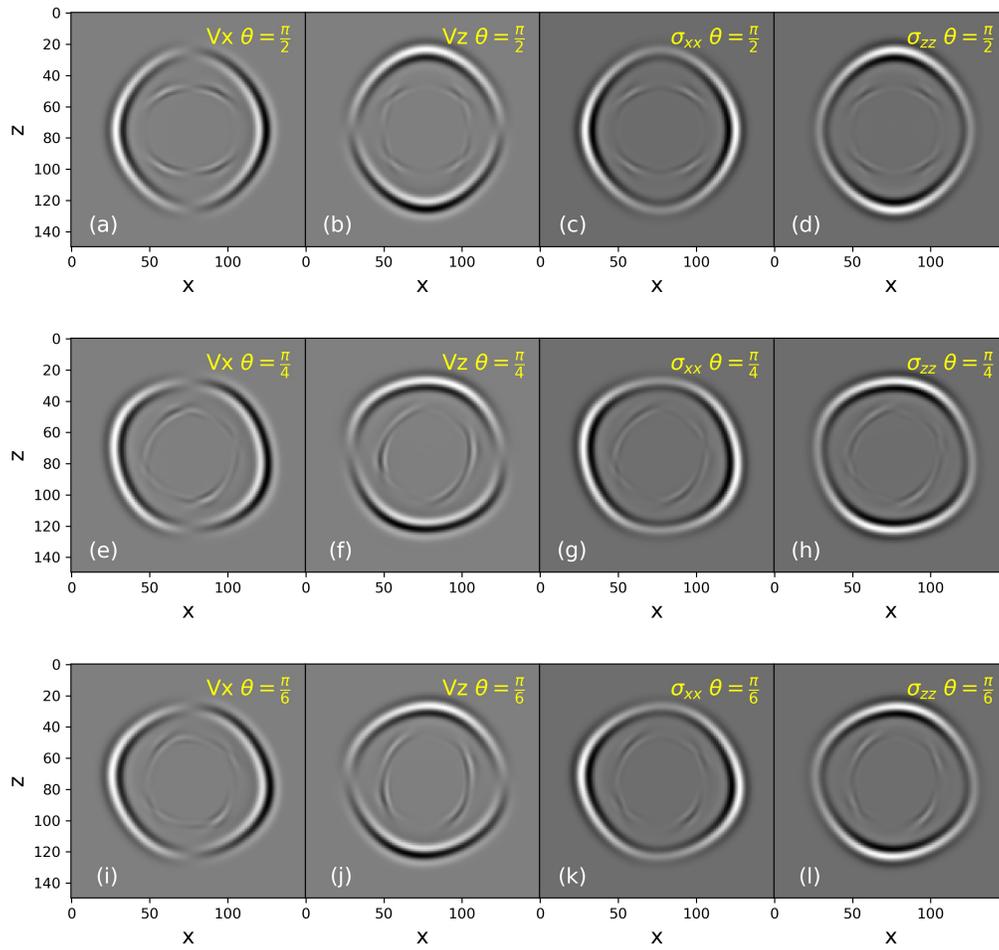


FIG. 1. TTI wave fields snapshot with different title angles. (a) V_x with $\theta = \frac{\pi}{2}$. (b) V_z with $\theta = \frac{\pi}{2}$. (c) σ_{xx} with $\theta = \frac{\pi}{2}$. (d) σ_{zz} with $\theta = \frac{\pi}{2}$. (e) V_x with $\theta = \frac{\pi}{4}$. (f) V_z with $\theta = \frac{\pi}{4}$. (g) σ_{xx} with $\theta = \frac{\pi}{4}$. (h) σ_{zz} with $\theta = \frac{\pi}{4}$. (i) V_x with $\theta = \frac{\pi}{6}$. (j) V_z with $\theta = \frac{\pi}{6}$. (k) σ_{xx} with $\theta = \frac{\pi}{6}$. (l) σ_{zz} with $\theta = \frac{\pi}{6}$.

1 (e)-(h) the title angle is $\theta = \frac{\pi}{4}$. Figure 1 (i)-(l) the title angle is $\theta = \frac{\pi}{6}$. The source of the wavelet is the Ricker's wavelet with main frequency 30Hz located in the center of the model. We can clearly see the corresponding change of the wavefields due to the change of the title angle. The wavefields are rotated according to the title angle.

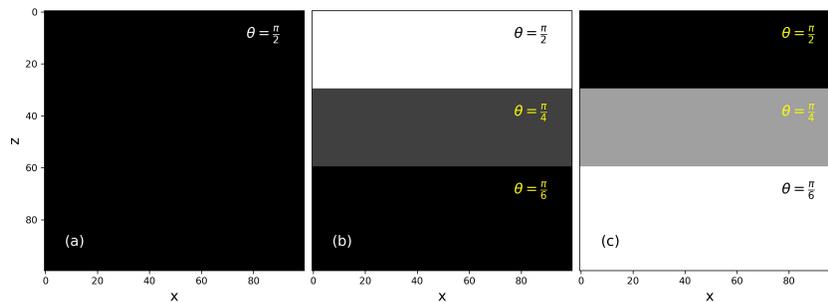


FIG. 2. TTI wave fields snapshot with different title angles. (a) title angle model $\theta = \frac{\pi}{2}$ (b) title angle layers model. $\theta_1 = \frac{\pi}{2}, \theta_2 = \frac{\pi}{4}, \theta_3 = \frac{\pi}{6}$. (c) (b) title angle layers model. $\theta_1 = -\frac{\pi}{2}, \theta_2 = -\frac{\pi}{4}, \theta_3 = -\frac{\pi}{6}$.

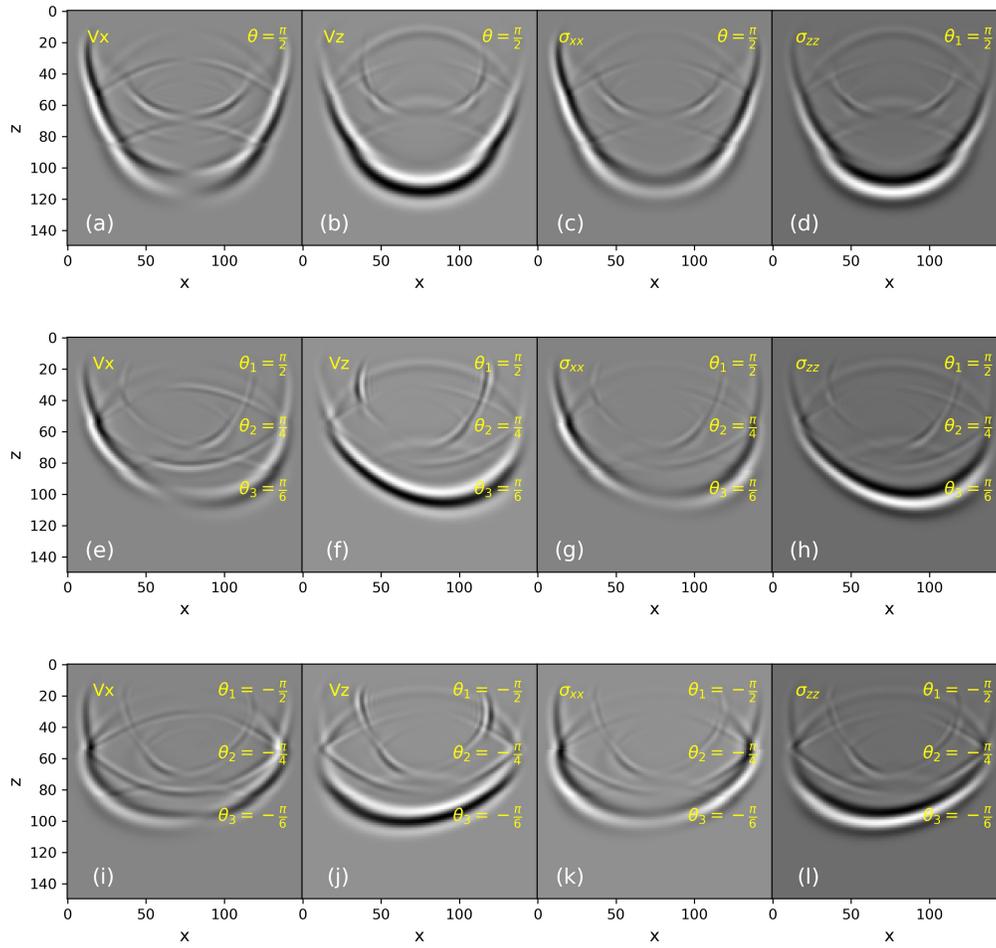


FIG. 3. TTI wave fields snapshot in layers model. (a) V_x with $\theta = \frac{\pi}{2}$. (b) V_z with $\theta = \frac{\pi}{2}$. (c) σ_{xx} with $\theta = \frac{\pi}{2}$. (d) σ_{zz} with $\theta = \frac{\pi}{2}$. (e) V_x with $\theta_1 = \frac{\pi}{2}$, $\theta_2 = \frac{\pi}{4}$, $\theta_3 = \frac{\pi}{6}$. (f) V_z with $\theta_1 = \frac{\pi}{2}$, $\theta_2 = \frac{\pi}{4}$, $\theta_3 = \frac{\pi}{6}$. (g) σ_{xx} with $\theta_1 = \frac{\pi}{2}$, $\theta_2 = \frac{\pi}{4}$, $\theta_3 = \frac{\pi}{6}$. (h) σ_{zz} with $\theta_1 = \frac{\pi}{2}$, $\theta_2 = \frac{\pi}{4}$, $\theta_3 = \frac{\pi}{6}$. (i) V_x with $\theta_1 = \frac{\pi}{2}$, $\theta_2 = \frac{\pi}{4}$, $\theta_3 = \frac{\pi}{6}$. (j) V_z with $\theta_1 = -\frac{\pi}{2}$, $\theta_2 = -\frac{\pi}{4}$, $\theta_3 = -\frac{\pi}{6}$. (k) σ_{xx} with $\theta_1 = -\frac{\pi}{2}$, $\theta_2 = -\frac{\pi}{4}$, $\theta_3 = -\frac{\pi}{6}$. (l) σ_{zz} with $\theta_1 = -\frac{\pi}{2}$, $\theta_2 = -\frac{\pi}{4}$, $\theta_3 = -\frac{\pi}{6}$.

Figure 2 shows another forward modeling test. In Figure 2 (a) the title angle is $\theta = \frac{\pi}{2}$, which means that this is a VTI model. The VTI model has three layers. Each layer has different values of stiffness modulus but the same title angle $\theta = \frac{\pi}{2}$. In Figure 2 (b), the title angle in first layer model is $\theta = \frac{\pi}{2}$, title angle in second layer has $\theta = \frac{\pi}{4}$, and title angle in third layer is $\theta = \frac{\pi}{6}$. In Figure 2 (c), the title angle in first layer model is $\theta = -\frac{\pi}{2}$, title angle in second layer has $\theta = -\frac{\pi}{4}$, and title angle in third layer is $\theta = -\frac{\pi}{6}$. Figure 3 also shows the wavefields, V_x , V_z , σ_{xx} , σ_{zz} in these TTI layers model.

Figure 3 (a)-(d) illustrate the wavefields calculated in model Figure 2 (a), which is the VTI model. Figure 3 (e)-(h) illustrate the wavefields calculated in title angle model Figure 2 (b), which is a TTI model with all positive title angles. Figure 3 (i)-(l) illustrate the wavefields calculated in title angle model, Figure 2 (c), which is a TTI model with all negative title angle. From the comparison between Figure 3 (e)-(h) and (i)-(l), we can see that with positive title angle the wavefields are drifted to the right, and with negative title

angles the wavefields are drifted to the left, which shows different waveform propagation pattern due to different title angle. Such a behavior, waveform changes due to the influence of the title angle, should be considered when we do full waveform inversion, since a very small change in the title angle can lead to the change of the waveform.

INVERSION WITH REGULARIZATION

Here we first introduce the elastic RNN misfits based on l_2 norm with high order TV regularization:

$$\begin{aligned} \Phi_{l_2}^{TV}(\mathbf{C}_{11}, \mathbf{C}_{13}, \mathbf{C}_{33}, \mathbf{C}_{44}, \theta, \alpha_1^{c11}, \alpha_1^{c13}, \alpha_1^{c33}, \alpha_1^{c44}, \alpha_1^\theta, \alpha_2^{c11}, \alpha_2^{c13}, \alpha_2^{c33}, \alpha_2^{c44}, \alpha_2^\theta) = \\ \frac{1}{2} \|\mathbf{D}_{syn}(\mathbf{C}_{11}, \mathbf{C}_{13}, \mathbf{C}_{33}, \mathbf{C}_{44}) - \mathbf{D}_{obs}\|_2^2 + \\ \alpha_1^{c11} \Theta_{TV}(\mathbf{C}_{11}) + \alpha_1^{c13} \Theta_{TV}(\mathbf{C}_{13}) + \alpha_1^{c33} \Theta_{TV}(\mathbf{C}_{33}) + \alpha_1^{c44} \Theta_{TV}(\mathbf{C}_{44}) + \alpha_1^\theta \Theta_{TV}(\theta) + \\ \alpha_2^{c11} \Upsilon_{TV}(\mathbf{C}_{11}) + \alpha_2^{c13} \Upsilon_{TV}(\mathbf{C}_{13}) + \alpha_2^{c33} \Upsilon_{TV}(\mathbf{C}_{33}) + \alpha_2^{c44} \Upsilon_{TV}(\mathbf{C}_{44}) + \alpha_2^\theta \Upsilon_{TV}(\theta) \end{aligned} \quad (15)$$

where $\alpha_1^{c11}, \alpha_1^{c13}, \alpha_1^{c33}, \alpha_1^{c44}, \alpha_1^\theta, \alpha_2^{c11}, \alpha_2^{c13}, \alpha_2^{c33}, \alpha_2^\theta$, are values of Lagrange multipliers. Θ_{TV} and Υ_{TV} represent first and second order TV regularization functions respectively. \mathbf{D}_{syn} represents the synthetic data, which is the function of the model parameters. Θ_{TV} and Υ_{TV} represent functions for calculating the first and second order TV regularization for the models.

The first order TV regularization term can be expressed as:

$$\begin{aligned} TV_1((\mathbf{m})) = \sum_{i=1}^{n-1} \sum_{j=1}^{m-1} |M_{i+1,j} - M_{i,j}| + \sum_{i=1}^{n-1} \sum_{j=1}^{m-1} |M_{i,j+1} - M_{i,j}| = \\ (\nabla_x, \nabla_z) \begin{pmatrix} \mathbf{m} \\ \mathbf{m} \end{pmatrix} = (\mathcal{L}_x, \mathcal{L}_z) \begin{pmatrix} \mathbf{m} \\ \mathbf{m} \end{pmatrix} = \Theta_{TV}(\mathbf{m}), \end{aligned} \quad (16)$$

where $M_{i,j}$ represents the element in model parameter vector. n and m are the grid numbers in x and z directions respectively. The second order TV regularization term can be expressed as:

$$\begin{aligned} TV_2((\mathbf{m})) = \sum_{i=1}^{n-1} \sum_{j=1}^{m-1} |M_{i+1,j} - 2M_{i,j} + M_{i-1,j}| + \sum_{i=1}^{n-1} \sum_{j=1}^{m-1} |M_{i,j+1} - 2M_{i,j} + M_{i,j-1}| \\ = (\nabla_{xx}, \nabla_{zz}) \begin{pmatrix} \mathbf{m} \\ \mathbf{m} \end{pmatrix} = (\mathcal{K}_{xx}, \mathcal{K}_{zz}) \begin{pmatrix} \mathbf{m} \\ \mathbf{m} \end{pmatrix} = \Upsilon_{TV}(\mathbf{m}). \end{aligned} \quad (17)$$

\mathcal{L}_x and \mathcal{L}_z are the first order differential vectors to calculate the first order total variations in x and z directions respectively. \mathcal{K}_{xx} and \mathcal{K}_{zz} are the second order differential vectors to calculate the second order total variations in x and z directions respectively.

The derivative of $\Phi_{l_2}^{TV}$ for each parameter, which is the gradient for \mathbf{C}_{11} , \mathbf{C}_{13} and \mathbf{C}_{33} ,

C_{44} based on the l_2^{TV} norm, can be expressed as:

$$\begin{pmatrix} \frac{\partial \Phi_{l_2}^{TV}}{\partial C_{11}} \\ \frac{\partial \Phi_{l_2}^{TV}}{\partial C_{13}} \\ \frac{\partial \Phi_{l_2}^{TV}}{\partial C_{33}} \\ \frac{\partial \Phi_{l_2}^{TV}}{\partial C_{44}} \\ \frac{\partial \Phi_{l_2}^{TV}}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \mathbf{G}_{12_{c11}} \\ \mathbf{G}_{12_{c13}} \\ \mathbf{G}_{12_{c33}} \\ \mathbf{G}_{12_{c44}} \\ \mathbf{G}_{12_{\theta}} \end{pmatrix} + \begin{pmatrix} \mathbf{R}_{12_{c11}} \\ \mathbf{R}_{12_{c13}} \\ \mathbf{R}_{12_{c33}} \\ \mathbf{R}_{12_{c44}} \\ \mathbf{R}_{12_{\theta}} \end{pmatrix}, \quad (18)$$

where $\mathbf{G}_{12_{c11}}, \mathbf{G}_{12_{c13}}, \mathbf{G}_{12_{c33}}, \mathbf{G}_{12_{c44}}, \mathbf{G}_{12_{\theta}}$ are the gradient vectors for parameters $C_{11}, C_{13}, C_{33}, C_{44}, \theta$ respectively. In this study we use the Automatic Differential method, to generate the gradients for the model parameters. The gradients for these parameters can be derived according to the adjoint state method as well.

Now we introduce how to choose the Lagrange multipliers. Let us first write the misfit function as:

$$\Phi^{TV} = J_D + J_{r1} + J_{r2}, \quad (19)$$

where J_D represents the any kind of norm misfit between observed data and synthetic data. $J_{r1} = \alpha_1^{c11} \Theta_{TV}(\mathbf{C}_{11}) + \alpha_1^{c13} \Theta_{TV}(\mathbf{C}_{13}) + \alpha_1^{c33} \Theta_{TV}(\mathbf{C}_{33}) + \alpha_1^{c44} \Theta_{TV}(\mathbf{C}_{44}) + \alpha_1^{\theta} \Theta_{TV}(\theta)$. $J_{r2} = \alpha_2^{c11} \Upsilon_{TV}(\mathbf{C}_{11}) + \alpha_2^{c13} \Upsilon_{TV}(\mathbf{C}_{13}) + \alpha_2^{c33} \Upsilon_{TV}(\mathbf{C}_{33}) + \alpha_2^{c44} \Upsilon_{TV}(\mathbf{C}_{44}) + \alpha_2^{\theta} \Upsilon_{TV}(\theta)$. The values $\alpha_1^{c11}, \alpha_1^{c13}, \alpha_1^{c33}, \alpha_1^{c44}, \alpha_1^{\theta}, \alpha_2^{c11}, \alpha_2^{c13}, \alpha_2^{c33}, \alpha_2^{\theta}$, are chosen according to the following formula:

$$K = \frac{J_D}{J_{r1} + J_{r2}}. \quad (20)$$

We should control the values for $\alpha_1^{c11}, \alpha_1^{c13}, \alpha_1^{c33}, \alpha_1^{c44}, \alpha_1^{\theta}, \alpha_2^{c11}, \alpha_2^{c13}, \alpha_2^{c33}, \alpha_2^{\theta}$ to keep the balance between the influence of the regulation terms and data misfit term. If K is too large, the data misfit J_D would dominate the final misfit value and the regularization term does not have enough influence on the inversion results. If K is too small, the regularization term would have too much influence on the inversion results. In this study, K value is chosen to between 0.1 and 1, to make sure the regularization terms have reasonable influence in the inversion. K should be relatively large when noise occurs in the data.

NUMERICAL TESTS

Figure 4 (a), (d),(g),(j),(m) are the true models for $C_{11}, C_{13}, C_{33}, C_{44}$, and θ respectively. Figure 4 (b), (e),(h),(k),(n) are the initial models for $C_{11}, C_{13}, C_{33}, C_{44}$, and θ respectively. In this numerical test we use the Ricker's wavelet as source. The main frequency of the source is 30Hz. $dx = dz = 7m$. The size of the model is 30×50 .

Figure 5 shows the gradients given in this TTI inversion method with high order total variation regularization. Figure 5 (a) is the gradient for C_{11} . Figure 5 (b) is the gradient for C_{13} . Figure 5 (c) is the gradient for C_{33} . Figure 5 (d) is the gradient for C_{44} . Figure 5 (e) is the gradient for θ . Figure 4 (c), (f),(i),(l),(o) are the inversion results for $C_{11}, C_{13}, C_{33}, C_{44}$, and θ respectively. Figure 6 shows the inversion profiles Profiles through the recovered elastic models. Figure 6 (a) C_{11} inversion at 150m of the model. Figure 6 (b) C_{13} inversion at 150m of the model. Figure 6 (c) C_{33} inversion at 150m of the model. Figure 6 (d) C_{44} inversion at 150m of the model. Figure 6 (e) θ inversion at 150m of the model. From the

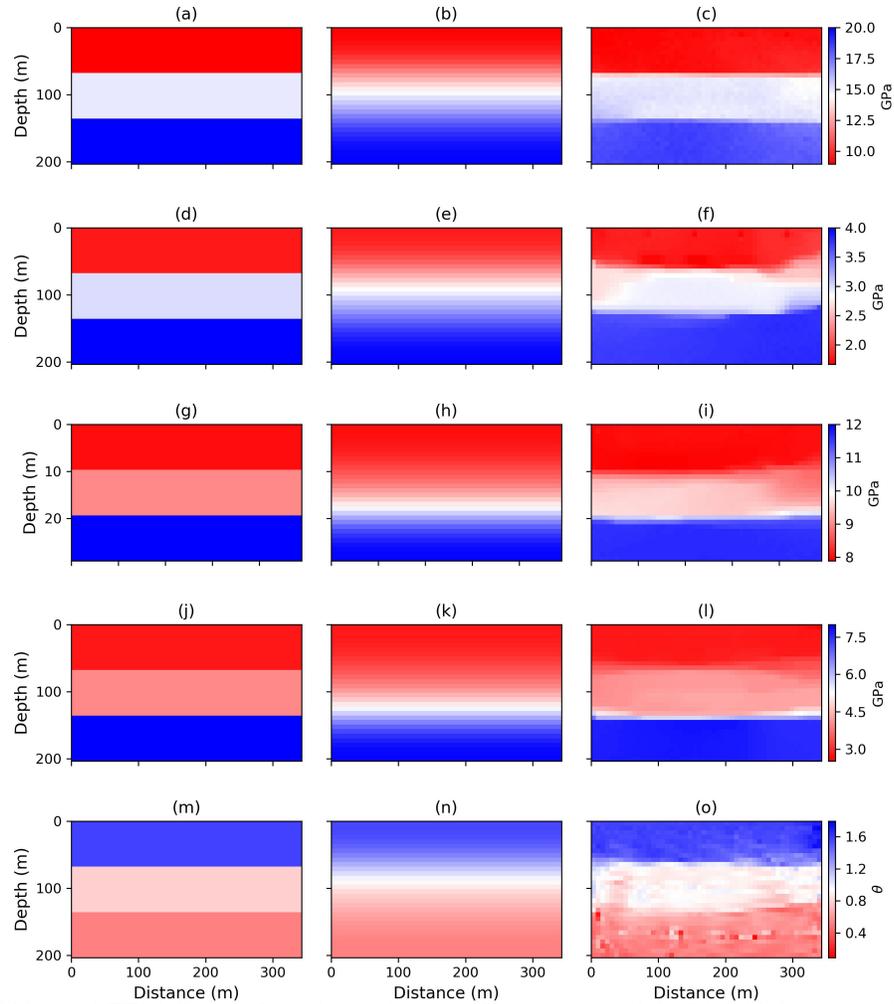


FIG. 4. RNN based TTI full waveform inversion. (a) True C_{11} model. (b) Initial C_{11} model. (c) C_{11} inversion result. (d) True C_{13} model. (e) Initial C_{13} model. (f) C_{13} inversion result. (g) True C_{33} model. (h) Initial C_{33} model. (i) C_{33} inversion result. (j) True C_{44} model. (k) Initial C_{44} model. (l) C_{44} inversion result. (m) True θ model. (n) Initial θ model. (o) θ inversion result.

comparison between the inversion results and the true models, we can see that the RNN based TTI inversion with high order TV regulation can give correct inversion results for the inversion. In Figure 6 (a), (b), (d) the inversions all gives the true right inversion results. However, in 6 (c) and (e) the inversion for C_{33} and θ , the layer is a little bit misplaced. This may due to the poor initial model which leads to the wrong position for the inversion results.

In this numerical test, we use part of the BP model to test the efficiency of this inversion method. The source of the wavelet is the Ricker's wavelet with a main frequency of 35Hz. The size of the model is 60×100 grid points. The grid length of the model is $dz = dx = 7m$. 7 shots are evenly distributed on the top of the model. Every grid point at the top of the model is located a receiver. Figure 7 (a), (c), (e), (g), (e) are the true C_{11} , C_{13} , C_{33} , C_{44} and θ models respectively. Figure 7 (b), (d), (f), (h), (f) are the inversion results for C_{11} , C_{13} , C_{33} , C_{44} and θ models respectively. From the comparison of the inversion with the true models we can see that the anticline structure of the model has been correctly updated,

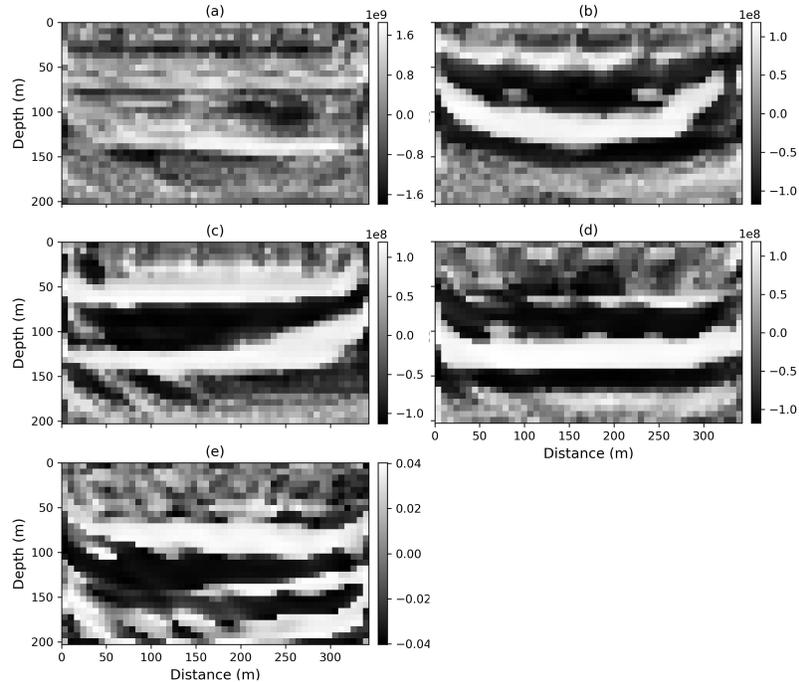


FIG. 5. Gradient for each parameters. (a) C_{11} gradient. (b) C_{13} gradient. (c) C_{33} gradient. (d) C_{44} gradient. (e) θ gradient.

however with some noise between the fine layers. This may be caused because of the high nonlinear relationship between the elastic stiffness parameters and the true angles. Figure 8 shows the profiles through the recovered elastic models at 100m of the models. Figure 8 (a) shows the inversion for C_{11} . Figure 8 (b) shows the inversion for C_{13} . Figure 8 (c) shows the inversion for C_{33} . Figure 8 (d) shows the inversion for C_{44} . Figure 8 (e) shows the inversion for θ . The red lines are the true values, the blue lines are the initial values and the green lines are inversion results. From Figure 8 we can see that C_{11} , C_{13} , C_{33} , C_{44} has been correctly updated. However the value for θ is very hard to recover in this model.

Figure 9 shows the gradients given by this method. Figure 9 (a) is the gradient for C_{11} . Figure 9 (b) is the gradient for C_{13} . Figure 9 (c) is the gradient for C_{33} . Figure 9 (d) is the gradient for C_{44} . Figure 9 (e) is the gradient for θ . During the forward propagation, how the synthetic records are calculated would be recorded which forms the Dynamic Computational Graph. The gradients of the trainable parameters would be calculated by using this Dynamic Computational Graph, according to the backpropagation method. After the gradients are calculated, with an optimization method and the step lengths for each model, we can get the directions to update the trainable parameters and reduce the misfit. With the Automatic Differentiation engine built in the machine learning library, the exact gradients would be calculated according to this Dynamic Computational Graph. Not only the wavefields would be saved in RAM, but the mathematical operation and internal variables would also be stored in the computer. Even though, the automatic differentiation method differs from the conventional adjoint state method. I myself believe its ability in extreme complex media when wave equations are no longer self-adjoint. Also, this part still needs to be further tested as well.

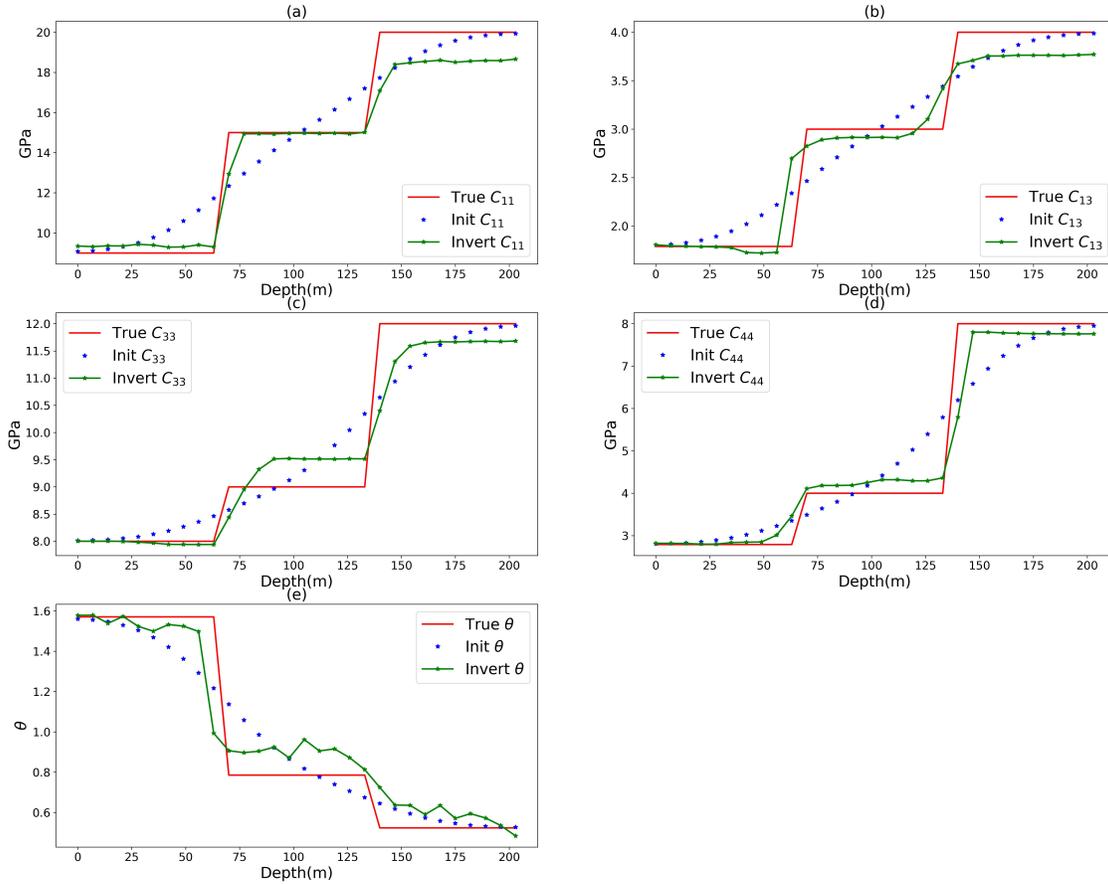


FIG. 6. Profiles through the recovered elastic models. (a) C_{11} inversion at 150m of the model. (b) C_{13} inversion at 150m of the model. (c) C_{33} inversion at 150m of the model. (d) C_{44} inversion at 150m of the model. (e) θ inversion at 150m of the model.

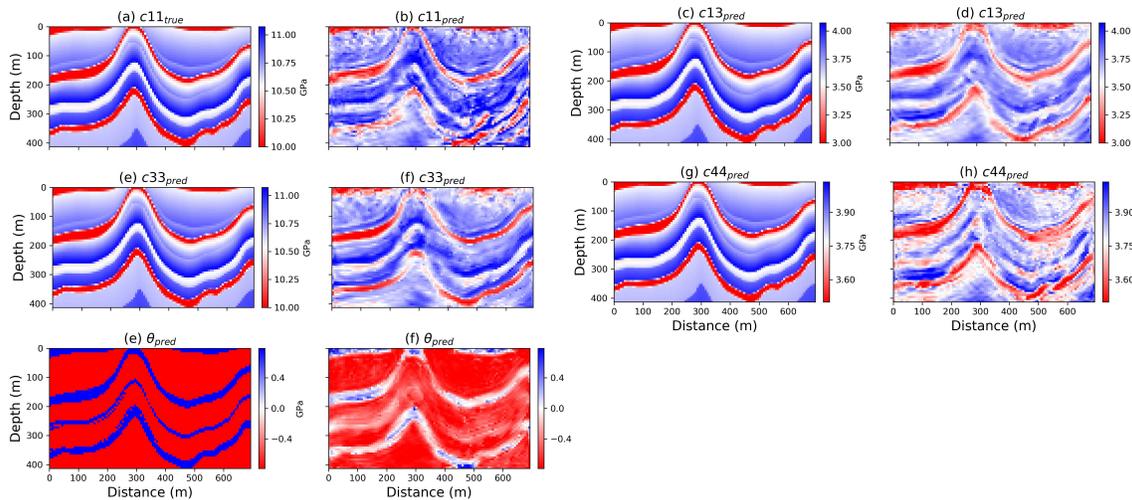


FIG. 7. Part of BP model inversion results. (a) True C_{11} . (b) RNN TTI FWI C_{11} inversion result. (c) True C_{13} . (d) RNN TTI FWI C_{11} inversion result. (e) True C_{33} . (f) RNN TTI FWI C_{11} inversion result. (g) True C_{44} . (h) RNN TTI FWI C_{11} inversion result. (i) True θ . (j) RNN TTI FWI C_{11} inversion result.

CONCLUSIONS AND FIGURE STUDY

In this study, based on the TTI staggered grid stress velocity wave equation, we build the TTI RNN cell and performed the TTI full waveform inversion, which forms the theory-

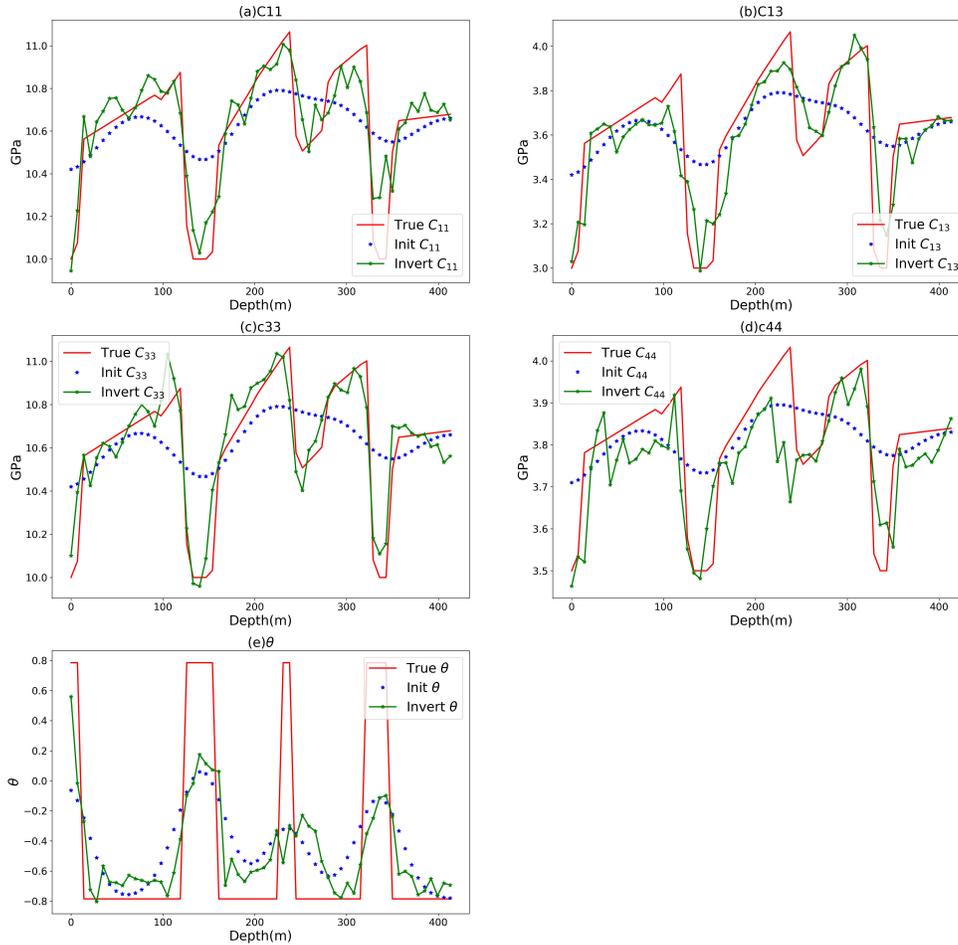


FIG. 8. Profiles through the recovered elastic models. (a) C_{11} inversion at 150m of the model. (b) C_{13} inversion at 150m of the model. (c) C_{33} inversion at 150m of the model. (d) C_{44} inversion at 150m of the model. (e) θ inversion at 150m of the model.

based machine learning TTI full waveform inversion. This inversion framework would generate the gradient automatically, and the gradients are the exact gradients based on the forward computational graph. The title angle of the complex TTI media can be directly inverted.

Numerical tests show that the inversion framework we introduced would generate the correct inversion results for C_{11} , C_{13} , C_{33} , C_{44} and θ . The numerical tests show that the inversion of θ the title angle is the hardest to recover. With the Automatic differential method, the gradient of the complex media parameters based on complex misfits can be calculated. The numerical tests also proved the validation of this method.

For future study, first, we would like to do some comparison between the conventional adjoint state method and the automatic differential method and prove it mathematically. Second, we would like to apply these methods to the real data sets to prove the validation in real seismic data. Third, we will seek methods that can reduce the computational cost of this method, Fourth, we will combine more data-driven methods to mitigate the modeling error problems.

ACKNOWLEDGMENTS

We thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors, NSERC (Natural Science and Engineering Research Council of Canada) through the grants CRDPJ 461179-13 and CRDPJ 543578-19. Partial funding also came from the Canada First Research Excellence Fund. Author is supported by China scholarship Counsel.

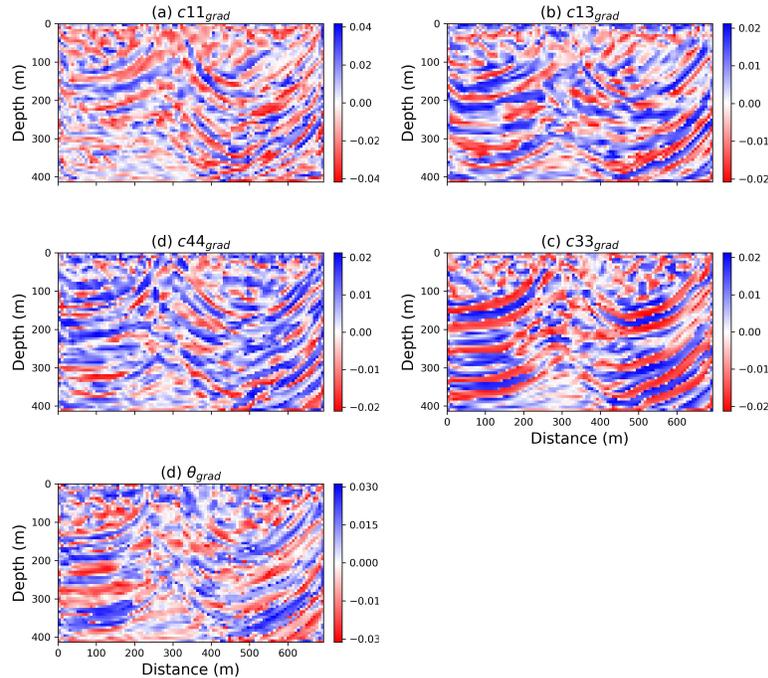


FIG. 9. Profiles through the recovered elastic models. (a) C_{11} Gradient. (b) C_{13} Gradient. (c) C_{33} Gradient. (d) C_{44} Gradient. (e) θ Gradient.

REFERENCES

- Cracknell, M. J., and Reading, A. M., 2013, The upside of uncertainty: Identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines: *Geophysics*, **78**, No. 3, WB113–WB126.
- Gholami, Y., Brossier, R., Operto, S., Prieux, V., Ribodetti, A., and Virieux, J., 2013, Which parameterization is suitable for acoustic vertical transverse isotropic full waveform inversion? part 2: Synthetic and real data case studies from valhall: *Geophysics*, **78**, No. 2, R107–R124.
- Hornby, B. E., Howie, J. M., and Ince, D. W., 2003, Anisotropy correction for deviated-well sonic logs: Application to seismic well tie: *Geophysics*, **68**, No. 2, 464–471.
- Hua, B., and McMechan, G. A., 2003, Parsimonious 2d prestack kirchhoff depth migration: *Geophysics*, **68**, No. 3, 1043–1051.
- Lin, Y., and Wu, Y., 2018, Inversionnet: A real-time and accurate full waveform inversion with convolutional neural network: *The Journal of the Acoustical Society of America*, **144**, No. 3, 1683–1683.
- Mosser, L., Dubrulle, O., and Blunt, M., 2018, Stochastic seismic waveform inversion using generative adversarial networks as a geological prior, *in* First EAGE/PESGB Workshop Machine Learning.
- Oropeza, E. V., and McMechan, G. A., 2014, Common-reflection-point migration velocity analysis of 2d p-wave data from tti media: *Geophysics*, **79**, No. 3, C65–C79.

- Reading, A. M., Cracknell, M. J., Bombardieri, D. J., and Chalke, T., 2015, Combining machine learning and geophysical inversion for applied geophysics: ASEG Extended Abstracts, **2015**, No. 1, 1–5.
- Sun, H., and Demanet, L., 2018, Low frequency extrapolation with deep learning, *in* SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 2011–2015.
- Sun, J., Niu, Z., Innanen, K. A., Li, J., and Trad, D. O., 2020, A theory-guided deep-learning formulation and optimization of seismic waveform inversion: *Geophysics*, **85**, No. 2, R87–R99.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, No. 8, 1259–1266.
- Wang, X., and Tsvankin, I., 2013, Ray-based gridded tomography for tilted transversely isotropic media: *Geophysics*, **78**, No. 1, C11–C23.
- Wrona, T., Pan, I., Gawthorpe, R. L., and Fossen, H., 2018, Seismic facies analysis using machine learning: *Geophysics*, **83**, No. 5, O83–O95.
- Yang, F., and Ma, J., 2019, Deep-learning inversion: a next generation seismic velocity-model building method: *Geophysics*, **84**, No. 4, 1–133.