Hindawi

Research Article

Geometrical-Feature-Preserving Adjoint Tomography of Near-Surface Structure with Seismic Early Arrival

Chao Zhang [1,2] **Cheng Li,**[3] **and Jian Cai**[3]

¹Key Laboratory of Marine Hazards Forecasting, Ministry of Natural Resources, Hohai University, Nanjing 210098, China
²College of Oceanography, Hohai University, Nanjing 210098, China
³China University of Science and Technology of China, Hefei, Anhui, China

Correspondence should be addressed to Chao Zhang; czhang18@hhu.edu.cn

Early arrival waveform inversion (EWI) is an essential approach to obtaining velocity structures in near-surface. Due to suffering from a cycle-skipping issue, it is difficult to reach the global minima for conventional EWI with the misfit function of least-squares norm (L2-norm). Following the optimal transportation theory, we developed an EWI solution with a new objective function based on quadratic-Wasserstein-metric (W2-norm) to maintain the geometric characteristics of the distribution and improve the stability and convexity of the inverse problem. First, we gave the continuous form of the adjoint source and the Frechet gradient of the Wasserstein metric for seismic early arrival, which leads to an easy and efficient way to implement in the adjoint-state method. Then, we conducted two synthetic experiments on the target model containing some velocity anomalies and hidden layers to test its effectiveness in mapping accurate and high-resolution near-surface velocity structure. The results show that the W2-normed EWI can mitigate cycle-skipping issues compared with the L2-normed EWI. In addition, it can deal with hidden layers and is robust in terms of noise. The application to a real dataset indicates that this new solution can recover more details in the shallow structure, especially in the aspect of dealing with hidden layers.

1. Introduction

Accurate near-surface solution is essential for mapping subsurface structures of the Earth in seismic industry. Generally, the irregular geology and undulating topography of the near-surface weathered layers lead to complex velocity variations [1]. If the structure in the near-surface region is not accurately recovered, the deeper reflections can be strongly degraded after migration [2]. First-arrival travel time tomography, which inverts the first-arrival travel times, has been widely applied in near-surface imaging [3, 4]. However, travel time tomography may be limited when applicating in geologically complex areas because it is limited by the assumption of high frequency approximation. Then, several methods have been developed to overcome the limitation, including the fat ray tomography method [5], the Fresnel-volume tomography method [6], and the wave-

equation tomography methods [7]. Full waveform inversion (FWI) [8–13] is another state-of-the-art imaging method used to overcome high-frequency limitation, which solves full wave equations to simulate accurately the propagation of seismic waves in complex media and considering full waveform information in inversion. FWI is able to yield better-resolved Earth models compared with conventional ray-based approaches.

For near-surface imaging, the early arrivals (always defined as the signals arriving within a few periods of the first arrival) rather than the full-wave information are the most commonly used to invert the near-surface structure at shallow depths (0–500 m) [14]. Generally, early arrivals could include seismic refractions, direct waves or shallow seismic reflections. Recently, several studies of early arrival waveform inversion (EWI) have been investigated. Sheng et al. [14] proposed an EWI method for imaging the near-

surface velocity, by minimizing the data misfit between the synthetic and observed waveform in the time window of early arrivals, and they later applied this method to marine data [15] and land data [16]. Shen [17] developed a weighted EWI method by matching the phase instead of amplitude during inversion, which was later applied to shallow land data [18]. Then, a joint inversion of early arrival waveform and reflection waveform was proposed to better reconstruct the shallow structures [19], by combining the low-wave-number and high-wavenumber in the near surface velocity model. Also, this method was applied to a real ocean bottom cable (OBC) case study with gas cloud [20].

However, the misfit function of both FWI and EWI is highly nonlinear, which makes it easy to trap in the local minima before reaching the global minima and results in poor convergence during inversion. That is, cycle-skipping problems usually exist in waveform inversion (Figure 1). To partly mitigate this problem, several methods have been developed in recent years. Ma and Hale [21] proposed the hierarchical strategy of waveform inversion, which firstly resolve the low-wavenumber content by wave-equation travel time tomography and then use FWI to constrain high-wavenumber details in the model. Wu et al. [22] developed an envelope inversion to mitigate cycle-skipping issues by extracting low-frequency component in seismic data. As a result, it can provide a good initial model with low-wave-number detail for the FWI. Bozdag et al. [23] developed more sophisticated objective functions to mitigate the local minima, e.g., using the instantaneous phase (IP) and amplitude envelope. Then, adaptive FWI methods were also proposed [24, 25], which reformulates the misfit functions by some matching filters and update velocity model by fixing time shifts between the synthetic and observed waveform. Zhang et al. [26] proposed waveform-based joint inversion of surface waves with traveltime and Z/H amplitude ratios to constrain shallow surface and provide an initial model for FWI.

However, most of the above studies used the misfit functions of least-squares norm (L2-norm). It is also well known that some obstacles exist in FWI based on L2-norm [27]. The main obstacles include the following: (i) L2-norm misfit function is always nonconvex in the respect of model parameters, making it easy to fall into local minima. (ii) The L2-norm cannot deal with noise in seismic data, which will be inverted into the final model and yield an incorrect result. If we examine the data misfit between synthetic waveform and observed waveform from another viewpoint, there is a mapping which may exist between these two datasets [27–29]. Thus, Engquist and Froese [30] proposed a novel misfit measurement which is associated with optimal transport. The squared Wasserstein metric (W2) used in this method is convex with respect to time shift leading to mitigate the cycle-skipping problems. The theory of optimal transport has been developed in the interdisciplinary and cutting-edge research fields of geometry, probability theory, and partial differential equations and has been applied in the fields of engineering and medicine in recent years [31, 32]. Yang et al. [27] applied the W2-norm to FWI in the field of seismic exploration and achieved good results. It should be
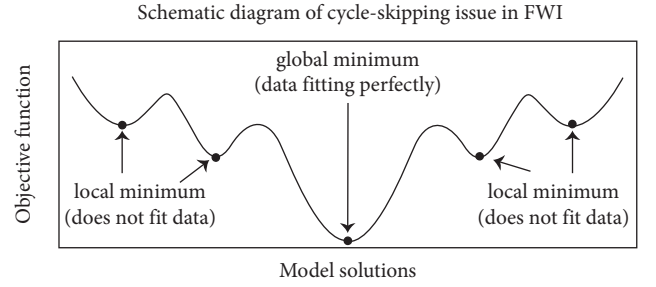
Schematic diagram of cycle-skipping issue in FWI

Figure 1: Schematic diagram of cycle-skipping issues in waveform inversion.

noted that no studies have used the quadratic-Wasserstein-metric in EWI solution. In addition, the geometric preserving property of the W2 metric has not been revealed in previous work.

In this paper, we used the quadratic Wasserstein metric as the misfit function in EWI. First, we gave the definition of the Wasserstein distance and derived the strict continuous form of the gradient direction of the target function under the W2 metric for early arrival. Then, we show the advantages of the quadratic Wasserstein distance by comparing the W2-norm-based misfit function with the L2-norm. We applied the newly developed seismic tomography method based on the W2-norm to EWI, and two numerical experiments were tested to show its effectiveness in mapping accurate and high-resolution near-surface velocity structure. Finally, the new method was tested using real data.

2. Theory and Methodology

The full waveform inversion can be considered as a PDE constrained optimization problem from the view of mathematics

$$\int c_T(x) = \arg\min_{c(x)} \sum_{i=1}^{N} \sum_{j=1}^{M} \chi_{ij}(c(x)). \tag{1}$$

Correspondingly, the misfit function χ_{ij} of the source-receiver pair in index (i,j) is defined as

$$\chi_{ij}(c(x)) = \mathcal{D}\big(s_{ij}(t; c(\mathbf{x})), d_{ij}(t)\big), \tag{2}$$

where N indicates the total number of seismic events and M indicates seismic signals for each event. D is the distance function that measures the difference between the synthetic waveform $s_{ij}(t; c(x))$ and observed waveform $d_{ij}(t)$.

2.1. Definition of the Wasserstein Distance. Assuming that there are two probability distributions μ and v, we identify a homeomorphic mapping onto itself, which needs to satisfy two conditions simultaneously: (i) measure-preserving mapping and (ii) minimum transmission cost. The mapping satisfying the above conditions is called the optimal transport mapping between these two probability distributions.

For any Borel set $\psi: \Omega \longrightarrow \Omega$, mapping S maps the probability distribution μ to v and satisfies

$$\int_\Omega \psi(y)\mathrm{d}v(y) = \int_\Omega \psi(S(x))\mathrm{d}\mu(x). \tag{3}$$

If the secondary transportation cost of moving a point from x to $S(x)$ is $|x - S(x)|^2$, the corresponding secondary transmission cost of the transportation mapping is defined as

$$E(S) := \int_\Omega |x - S(x)|^2 \mathrm{d}\mu(x). \tag{4}$$

In all of the sets of the self-mapping preserving measure, the mapping with the minimum transmission cost is called the optimal transmission mapping.

$$S^* = \underset{S^*\mu=v}{\arg\min}\, E(S). \tag{5}$$

Wasserstein distance $W^2(\mu, v)$ is defined as the transmission cost of the optimal transportation map [31], which is expressed as follows:

$$W^2(\mu, v) = \inf_{S_{u,v}\in M} \int_\Omega |x - S_{\mu,v}(x)|^2 \mu(x)\mathrm{d}x. \tag{6}$$

Brenier [33] proved that a convex function $u: \Omega \longrightarrow \sim^+$ exists, and its gradient mapping, namely, $S: x\mapsto\nabla u(x)$, is the unique mapping of the optimal transportation.

f and g are two probability density functions on Ω, satisfying $\mu = f\mathrm{d}x$ and $v = g\mathrm{d}x$. If μ retains absolute continuity in the sense of the Lebasque measure, w, then (i) a unique optimal transmission map S exists and (ii) a convex function u: $\Omega \longrightarrow \Omega$, exists; so, the optimal transmission map satisfies $S = \nabla u$.

If the above conclusions are substituted into equation (3), we obtain the following equation for any $\psi \in C_c^\infty(\Omega)$:

$$\int_\Omega \psi(S(x))f(x)\mathrm{d}x = \int_\Omega \psi(y)g(y)\mathrm{d}y$$
$$= \int_\Omega \psi(S(x))g(S(x))|\det\nabla S(x)|\mathrm{d}x. \tag{7}$$

Thus,

$$f(x) = g(S(x))|\det\nabla S(x)|. \tag{8}$$

Under the condition $S = \nabla u$, μ satisfies the Monge–Ampère equation.

$$\det D^2 u = \frac{f}{g \circ \nabla u}. \tag{9}$$

The Monge–Ampère equation was proposed by Monge [34] and Ampère [35] and has been widely used in the optimal transport theory and astrophysics [33]. The information presented above indicates that we must solve the Monge–Ampère equation to obtain the optimal transmission map. However, the high-dimensional Monge–Ampère equation has no analytical solution, and the numerical solution is also very complex. For the one-dimensional case, the

traditional histogram equalization is the optimal transmission mapping [36, 37]. If f and g are two continuous 1-D probability distributions, their corresponding cumulative distribution function (CDF) is denoted as

$$F(x) = \int_{-\infty}^x \mathrm{d}u,$$
$$G(y) = \int_{-\infty}^y \mathrm{d}v. \tag{10}$$

The optimal map is as follows:

$$S(x) := G^{-1} \circ F(x). \tag{11}$$

Assume that there is a seismogram with the length of T, and the 1-D quadratic Wasserstein distance $W^2(f, g)$ can be expressed as follows [27]:

$$W^2(f, g) = \int_0^T |t - G^{-1}(F(t))|^2 f(t)\mathrm{d}t. \tag{12}$$

2.2. Adjoint Source Based on the W2-Norm.

EWI is a partial differential equation constrained optimization problem, which updates the model by a gradient-based optimization method. Following the adjoint-state method, the Fréchet kernels are always obtained using the time integration of the interaction between the forward and adjoint wavefield. As we know, the Fréchet kernels determine the direction of model updating, while the adjoint source is the premise of Fréchet kernels calculation. Therefore, the derivation of adjoint source and gradient is critical to make it easy to implement in the adjoint-state method. Based on the early arrival information, we gave the strict continuous form of the gradient direction of the target functional under the W2 metric. The continuous form of the adjoint source can be seen as follows:

$$\frac{\delta W^2[f(t)]}{\delta f(t)} = \left(t - G^{-1}(F(t))\right)^2$$
$$- 2\int_t^{T_0} \left(t - G^{-1}(F(\varepsilon))\right)\frac{f(\varepsilon)}{g\left(G^{-1} \circ F(\varepsilon)\right)}\mathrm{d}\varepsilon, \tag{13}$$

where $g(t)$ and $f(t)$ are the synthetic and observed data, respectively, $h(t)$ is the perturbation of the theoretical waveform, and ε is a small constant.

2.3. Advantages of the Quadratic Wasserstein Distance.

There are a few of advantages for W2-norm compared with the L2-norm from a mathematical point view. The main advantages include (1) the convexity with respect to time shift and amplitude change benefit from the feature of geometrical preserving; (2) the insensitivity to noise. We set up two simple numerical examples to illustrate it.

First, we assumed that the observed dataset f is a Ricker wavelet and the synthetic dataset g has a time shift compared with f (Figure 2(a)). The dominant frequency of both
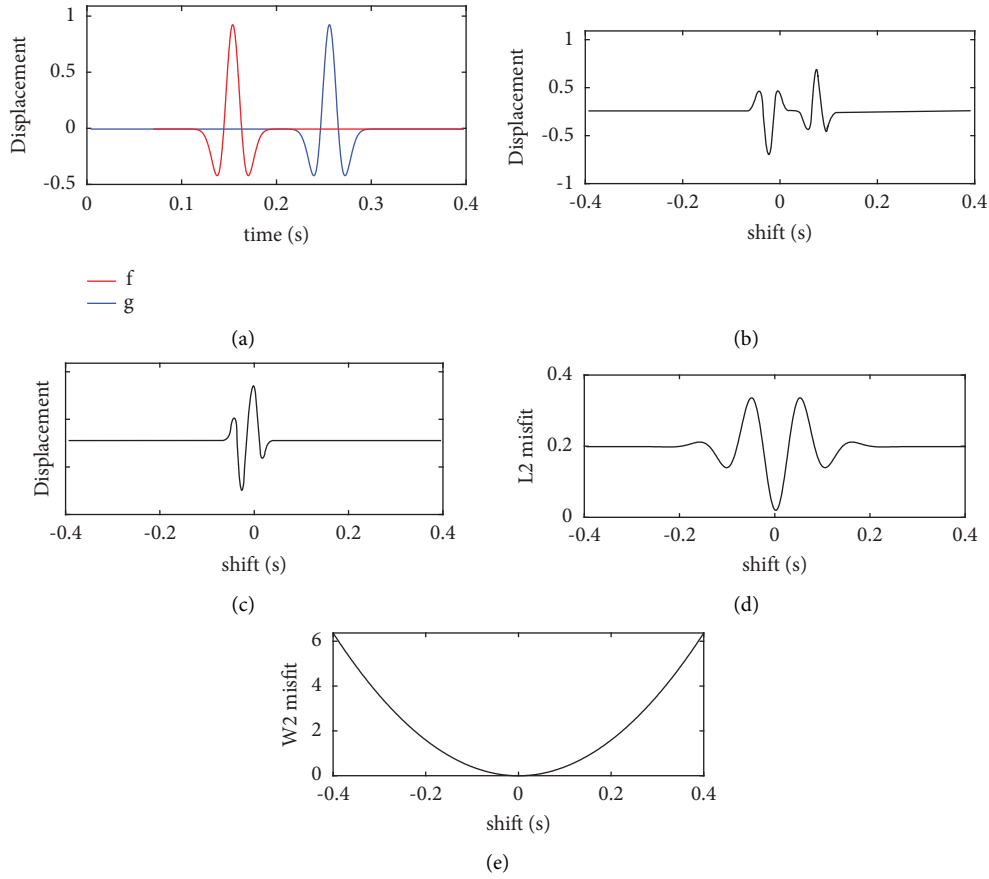
FIGURE 2: Comparison of the behavior of adjoint source and objective function for L2-norm and W2-norm. (a) Synthetic waveform: g and observed waveform: f, (b) the adjoint source of L2-norm, (c) the adjoint source of W2-norm, (d) L2-norm misfit, and (e) W2-norm misfit.
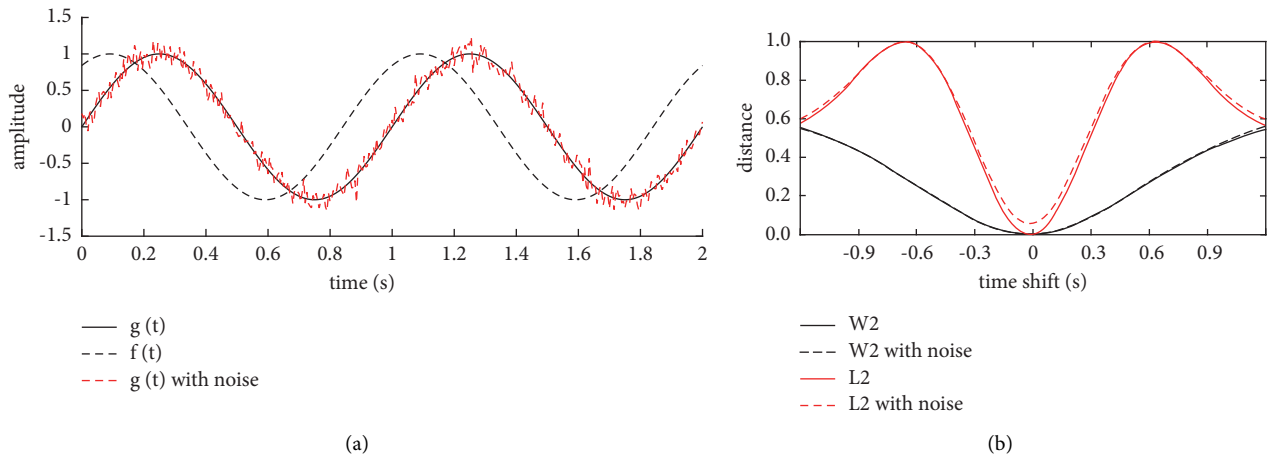

FIGURE 3: Comparison of the insensitivity to noise of the L2-norm and W2-norm. (a) Synthetic waveform: ft, observed waveform: gt, and observed data plus uniform random noise: gt + noise; (b) the distance between ft and gt and between ft and gt + noise under L2-norm and W2-norm.

signals is set as 10 Hz, and the time delay between f and g was set as greater than the half-period signal. It can be seen that there are 2 separate signal events existing in the adjoint source of L2-norm (Figure 2(b)). This will yield an incorrect Fréchet kernels calculated by correlating it with
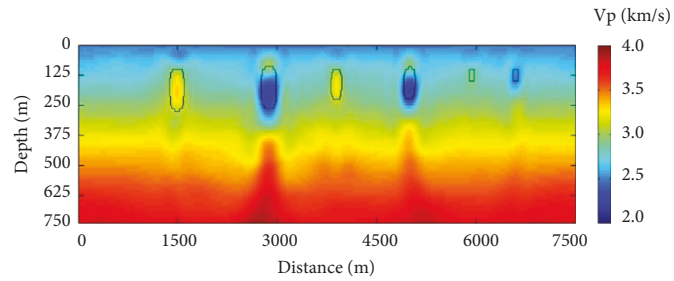
forward wavefield. The L2-norm misfit as a function of the shift between f and g is plotted in Figure 2(d), which shows that two local minima distribute near the global minima. This behavior may make it difficult for waveform inversion to converge towards global minima when the initial model
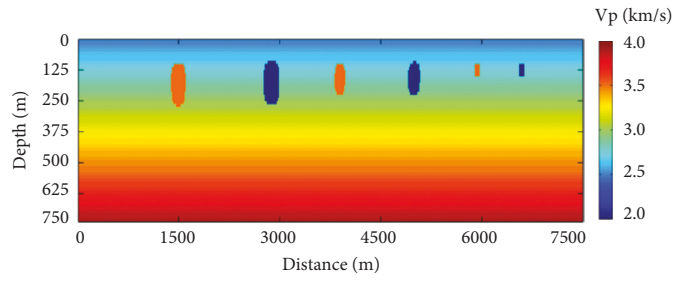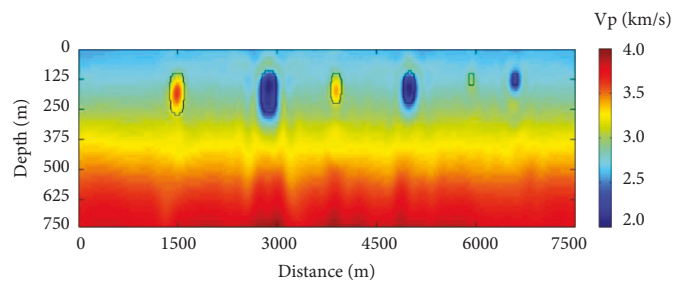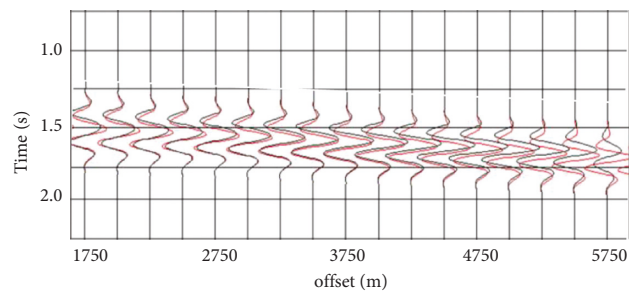
(a)
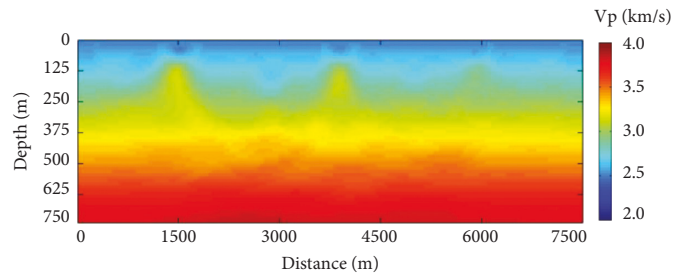
(b)
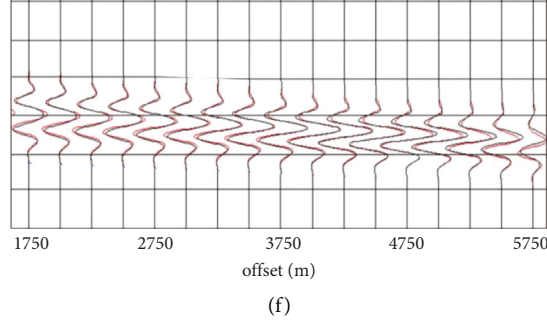
(c)

(d)

(e)

FIGURE 4: Continued.

(f)

FIGURE 4: The synthetic tests: (a) target model with alternating low and high velocity anomalies; (b) the travel time tomography results; (c) the inverted model of the L2-norm EWI; (d) the inverted model of the W2-norm EWI; (e) waveform fitting of early arrival between the synthetic data (red) and observed data (black) for the L2-norm EWI; (f) waveforms fitting of early arrival between the synthetic data (red) and observed data (black) for the W2-norm EWI.

is inaccurate. In contrast, the adjoint source of W2-norm has only a single event (Figure 2(c)), and their W2 distance function is a strictly convex function (Figure 2(e)). Therefore, when the gradient-based method is used to obtain the optimal solution, it is easy to trap in the local minima for L2-norm, but it can accurately converge to the global minimum for W2-norm.

For 1-D signals f and g with N data points, Yang et al. [27] proved that the influence of a uniform noise distribution $U[-c, c]$ on the L2 distance is $\mathcal{O}(1)$. L2 distance is largely affected by noise regardless of the number of points, whereas its effect on the W2 distance is $\mathcal{O}(1/\mathcal{N})$. Thus, if there are enough number of data points, the influence of noise on misfit function is negligible even if the noise is very strong. Here, we provide another simple numerical example to compare the difference in dealing with noise for the L2-norm and W2-norm. Assuming that there are two one-dimensional waveform signals, representing synthetic waveform ft and observed waveform gt, we added uniform random noise to the observed waveform gt to obtain gt + noise (Figure 3(a)). We calculated the distance between ft and gt and between ft and gt + noise using the L2-norm and W2-norm. The results shown in Figure 3(b) indicate that the noise has a large effect on the L2 distance, but it has little effect on the W2 distance.

2.4. Quadratic-Wasserstein-Metric-Based Early Arrival Waveform Inversion.
Early arrivals are defined as signals arriving after the first arrival within a few periods, which can be selected using a time window $[T_{\text{start.}}, T_{\text{end}}]$. For a single full waveform $f(t)$, the individual misfit function χ_i based on the W2 distance for early arrivals can be expressed as

$$\chi_i = W^2(f(t), g(t)) = \int_{T_{\text{start}}}^{T_{\text{end}}} \left| t - G^{-1}(F(t)) \right|^2 f(t) \mathrm{d}t, \quad (14)$$

where T_{start} and T_{end} are the start and end of the time window, respectively. Following the adjoint-state method, the perturbation of individual misfit $\delta\chi$ is linearly related to perturbations of P-wave velocity α, perturbations of S-wave velocity β, and perturbations of density ρ:
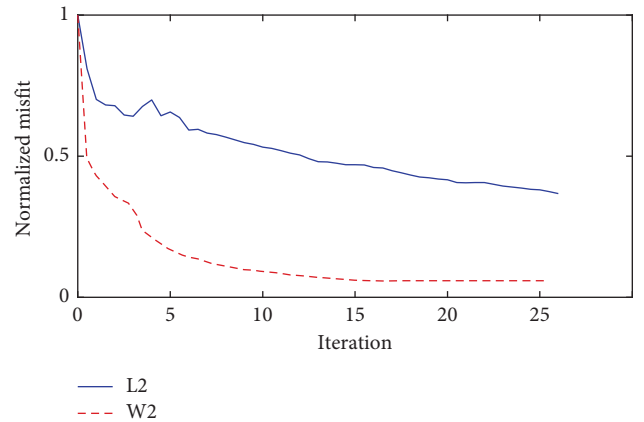

FIGURE 5: Reduction of waveform residuals with increasing number of iterations.

$$\delta\chi = \int \left(K_\alpha \delta \ln \alpha + K_\beta \delta \ln \beta + K_\rho \delta \ln \rho \right) \mathrm{d}^2 x, \quad (15)$$

where K_α, K_β, and K_ρ are the corresponding kernels for these model parameters (α, β, and ρ). The 2-D spectral element method (specfem2d) was used for simulating seismic wave propagation and to calculate the Fréchet derivatives numerically. The FLEXWIN open-source package [38] was used to select the time window for the early arrivals, in which the W2-based adjoint sources and misfits were calculated. The overall misfit function $\chi(m)$ for all of the selected windows is

$$\delta(m) = \frac{1}{N_\omega} \sum_{e=1}^{E} \sum_{i=1}^{N_\omega^e} \sum_{T_{i,\text{start}}}^{T_{i,\text{end}}} \left| t - G^{-1}(F(t)) \right|^2 f(t) \mathrm{d}t, \quad (16)$$

where m is the current Earth model, N_ω^e is the number of time windows of source e, E is the total number of sources, and N_ω is all of the selected time windows for the early arrivals. We use equation (13) to calculate the adjoint source in the corresponding time window, and Fréchet kernels are calculated by the interaction between the forward and adjoint wavefield. Then, all of Fréchet kernels were summed up to get the gradient, indicating
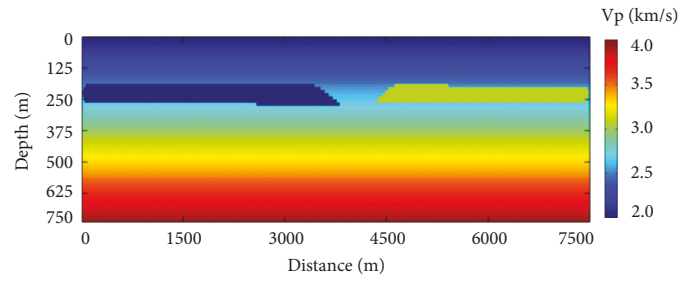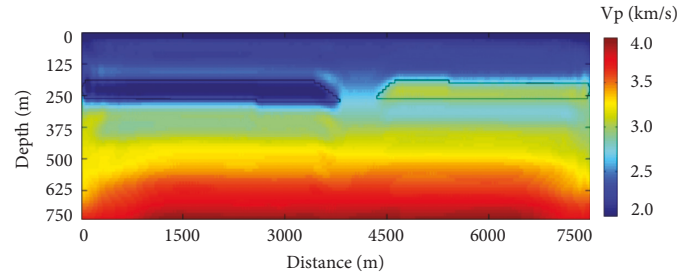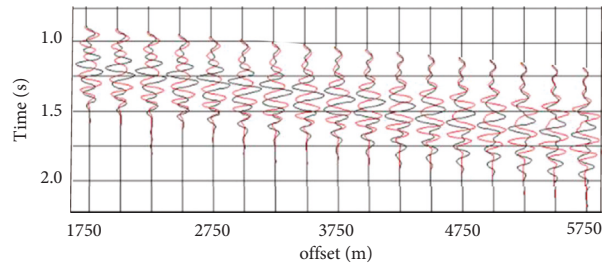
(a)
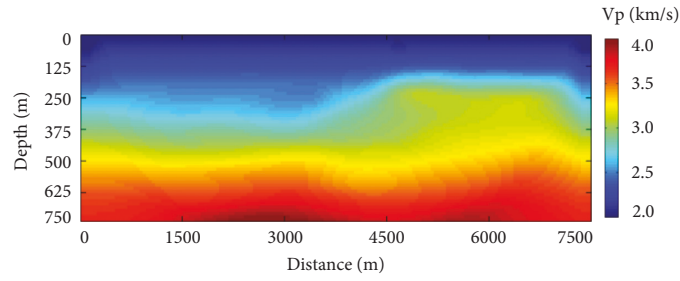
(b)
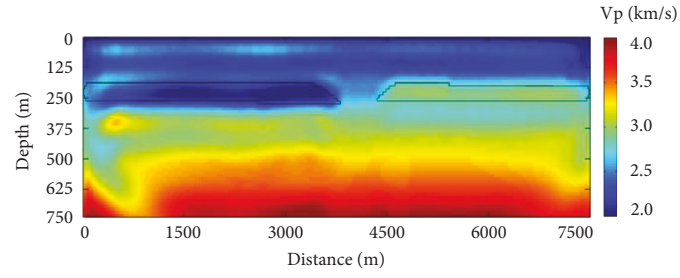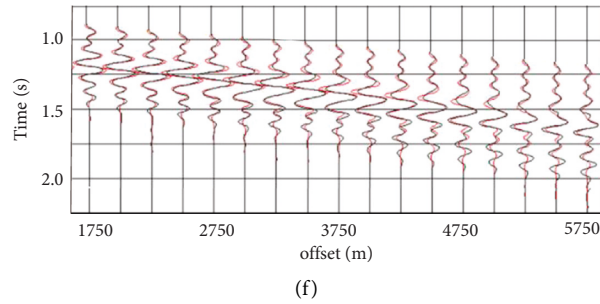
(c)

(d)

(e)

FIGURE 6: Continued.

(f)

FIGURE 6: The synthetic tests: (a) the target model with a hidden layer; (b) the travel time tomography results; (c) the inverted model of the L2-norm EWI; (d) the inverted model of the W2-norm EWI; (e) waveforms fitting of early arrival between the synthetic data (red) and observed data (black) for the L2-norm EWI; and (f) waveforms fitting of early arrival between the synthetic data (red) and observed data (black) for the W2-norm EWI.
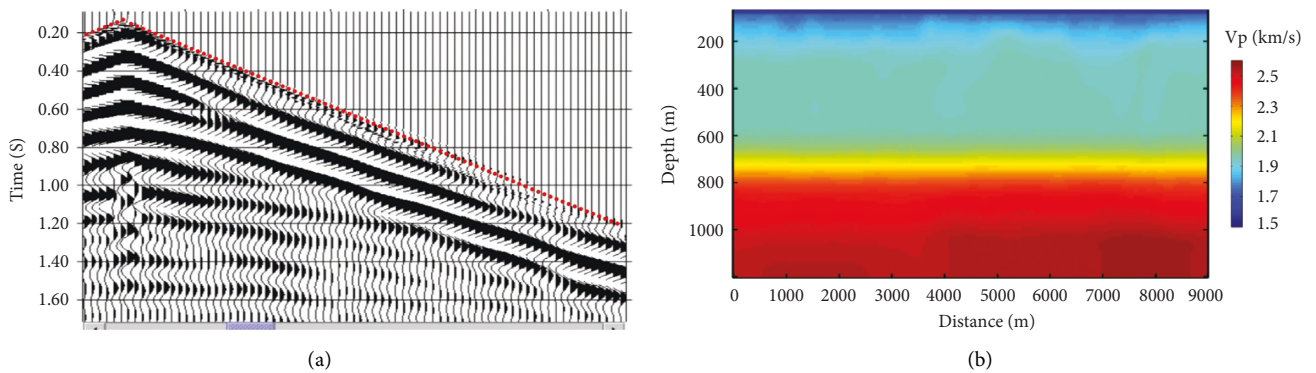

(a)

(b)

FIGURE 7: (a) CSG 50 from the marine seismic data: red dots denote the picked first-arrival travel times. (b) The inverted model of first-arrival travel time tomography.

(a)

(b)
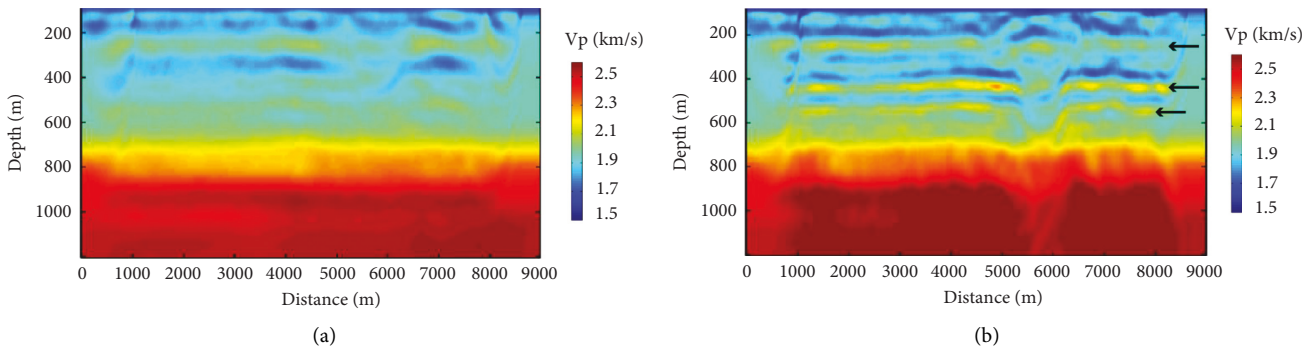
FIGURE 8: (a) The inverted model of the L2-norm EWI and (b) the inverted model of the W2-norm EWI.

the direction of EWI. The L-BFGS method [39] was used for the model updating, and the inversion would be terminated once the misfit decreases not quantifiably obvious.

3. Synthetic Experiments

To confirm the application of the W2-norm EWI, we tested the algorithm using two synthetic tests in which the model has typical features of near-surface structure.

3.1. Model with Velocity Anomalies. We set up a target model (Figure 4(a)) with six alternating high- and low-velocity anomalies. The speed of the low velocity anomalies was 2000 m/s, and the speed of the high velocity anomalies was 3500 m/s. The smallest scale of the scatterer is about 30 m, and the largest is close to 150 m. One hundred sources with a 75 m spacing and 250 receivers with a 30 m spacing were located from 0 to 7500 m along the surface of the target model. We also used a Ricker wavelet as the source time function. The background velocity model after excluding the velocity anomalies was regarded as the initial model for the

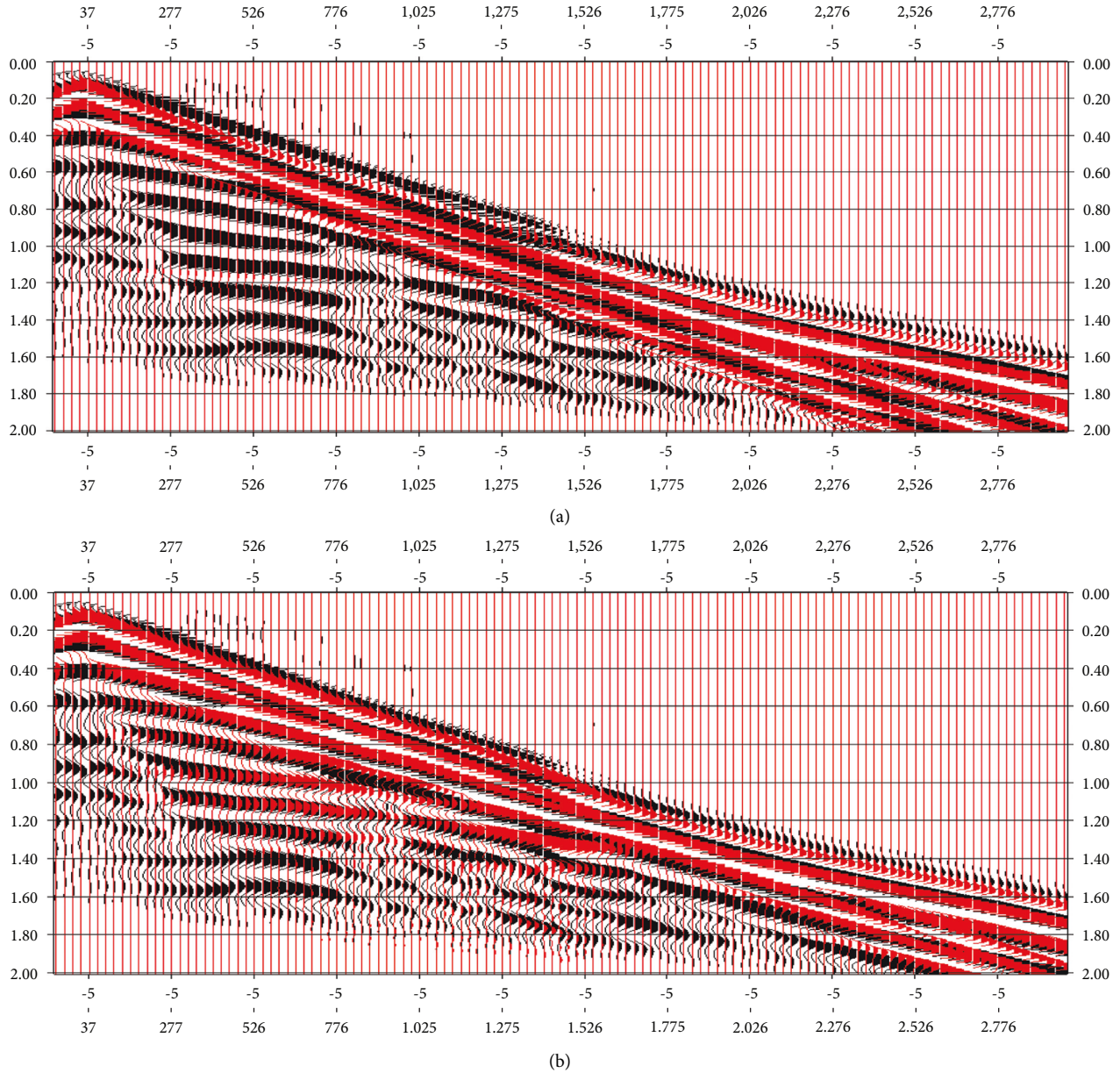FIGURE 9: (a) Waveform fitting of early arrival between the synthetic data (red) and observed data (black) for the L2-norm EWI; (b) waveform fitting of early arrival between the synthetic data (red) and observed data (black) for the W2-norm EWI.

travel time tomography. The results are shown in Figure 4(b). The results show that a smoothed model was attained after 10 iterations, which was used as the initial model for the L2-norm EWI and W2-norm EWI and was used to recover the low-wavenumber detail in shallow structures.

Figure 4(c) shows the L2-norm EWI solution. It can be seen that the L2-norm EWI clearly constrains the contours of the low and high velocity anomalies, and their positions are also consistent with those in the target model. However, the inverted anomalies are smaller than the target anomalies. We also found that there was a high-velocity artifact below the low-velocity anomaly. For the W2-norm EWI, the solution (Figure 4(d)) not only recovered the clear structures of

the low and high velocity anomalies but also constrained their positions close to the target model. In addition, the W2-norm EWI results contain fewer artifacts at the bottom of the low-velocity anomalies than L2-norm EWI. This suggests that the W2-norm EWI helps to tightly recover the near-surface velocity structures. Figures 4(e) and 4(f) display the corresponding waveform fitting between the observed waveform (black) and the synthetic waveform (red) associated with the L2-norm EWI and the W2-norm EWI at a shot gather. Figure 4(e) shows that it is reasonable for the overall waveform fitting, while the synthetic waveform does not actually match the observed waveform at the far offsets (4000–5750 m). Figure 4(f) shows the corresponding waveform match after the W2-norm EWI for the same shot

gather. The waveform fitting is obviously improved, especially for the seismogram at far offsets.

Figure 5 shows the normalized misfit over iterations of the L2-norm EWI and W2-norm EWI. It shows that the W2-norm EWI (dashed line) converges faster than the L2-norm EWI (solid line). The result indicates that W2-norm EWI has fast convergence speed, which can greatly improve the computational efficiency of waveform tomography.

3.2. Model with Hidden Layers. In seismic exploration, hidden layers cannot be solved by refraction travel time tomography [40]. This is why EWI has become important for dealing with complex near-surface problems. Here, we set up a synthetic test of a model with a hidden layer to confirm the capability of the W2-norm EWI in terms of solving hidden-layer problems. Figure 6(a) shows the target model, which contains a low velocity layer (2000 m/s) and a high velocity layer (3200 m/s). We chose the same geometry set up used in Section 3.1, and Ricker wavelet was set as the source wavelet. The background velocity model after excluding the hidden-layer was used as the initial model for the travel time tomography. As we expected, the first-arrival tomography failed to solve the low velocity layer (known as the hidden-layer) in the target model, producing a smooth result (Figure 6(b)), which was used as the initial model for the L2-norm EWI and the W2-norm EWI. As shown in Figure 6(c), the L2-norm EWI constrained the layers with clear contours, but a high-velocity artifact was generated above the layer. In addition, the low-velocity layer was bent instead of being a straight horizontal layer in the inverted model. Several high-velocity artifacts were also produced below the low-velocity layer. In comparison, the W2-norm EWI solution (Figure 6(d)) can successfully recover the clear outline both the high-velocity and low-velocity layers and constrain the layers to the accurate position. Also, there is almost no high-velocity artifacts observed above and below the layer. Figures 6(e) and 6(f) show the corresponding waveform fitting between the observed waveform (black) and the synthetic waveform (red) from a same common shot gather associated with the L2-norm EWI and W2-norm EWI, respectively. Because of suffering from cycle-skipping problem, a bad waveform match is seen in Figure 6(e), which indicates that the result of L2-norm EWI is still far from the true model. Figure 6(f) shows that the cycle-skipped data were mitigated when using the W2-norm EWI. This suggests that W2-norm EWI can avoid the cycle-skipping problem as well as providing a more reliable near-surface solution when hidden layers exist.

As was discussed in Section 3.1, one of the advantages of the Wasserstein distance is its capability to handle noise. To confirm the insensitivity of the W2-norm EWI to noise, we repeated the previous inversion using the data by adding uniform random noise. For the target model in Figures 4(a) and 6(a), we generated the synthetic waveform using specfem2d. Then, random noise with the signal-to-noise ratio (−4 dB) was added to the data. The other settings were the same as in the previous numerical experiment. Figures S1(a) and S1(c) show the results of the L2-norm EWI and W2-
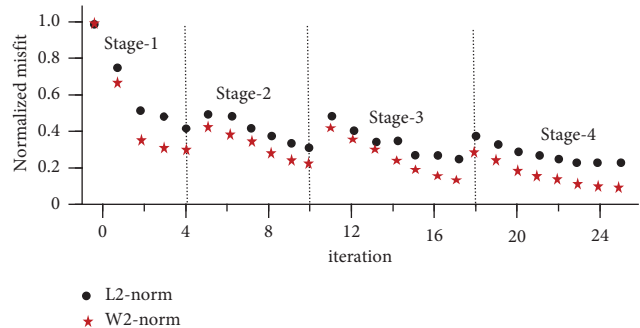

FIGURE 10: The normalized waveform misfit of L2-norm EWI and W2-norm EWI results.

TABLE 1: The applications in field data: relative misfit function of L2- and W2-norm EWI, respectively.

Iteration number	Relative misfit function	
	L2-norm EWI	W2-norm EWI
1	7.32×10^{-3}	6.50×10^{-3}
4	5.47×10^{-3}	3.32×10^{-3}
10	7.83×10^{-4}	5.46×10^{-4}
18	4.33×10^{-4}	2.15×10^{-4}
25	1.74×10^{-4}	1.02×10^{-4}

norm EWI with uniform random noise for the synthetic test on the target model with anomalies, respectively. The W2-norm EWI has a better solution for the velocity anomalies than the L2-norm EWI after the addition of noise. Similarly, the inverted model for the W2-norm EWI (Figure S1(d)) also constrains the shapes and locations of both the high- and low-velocity layers better than the L2-norm EWI (Figure S1(b)). This suggests that the W2-norm EWI can still converge reasonably well even when the noise is strong.

4. Applications in Field Data

We applied the W2-norm EWI to a real dataset from China. The dataset consists of 210 shots with a 40 m spacing and 450 receivers with a 20 m spacing. Figure 7(a) shows a typical shot gather (from shot 50). We manually picked the first-arrival travel times for the first-arrival travel time tomography. The results in Figure 7(b) show a rather smoothed velocity model, which can be used as a proper initial model for the EWI. Then, we muted the data with an early arrival time window using the FLEXWIN package [38] and conducted trace normalization [14]. To reduce the nonlinearity of the waveform misfit function, we adopted a hierarchical strategy, in which we successively inverted through 4 stages from low to high frequency: 10 Hz (stage 1), 20 Hz (stage 2), 30 Hz (stage 3), and 40 Hz (stage 4).

We performed two workflows: (1) the L2-norm EWI using the travel time tomography result as the input model; and (2) the W2-norm EWI using the travel time tomography results as the input model. We ran the same iterations for both methods. Overall, the L2-norm EWI (Figure 8(a)) and W2-norm EWI (Figure 8(b)) produced more velocity details than the travel time tomography alone. It can be seen that

the L2-norm EWI produced several high-velocity layers (e.g., at depths of ~200 m and 280 m) that do not exist in the travel time tomography solution. However, the interface of these layers also seems smoothed, and the low-velocity layers between the high-velocity layers are not obvious. In contrast, the W2-norm EWI shows more velocity details than the L2-norm EWI. In particular, several low-velocity layers/hidden layers were recovered among the high-velocity layers (highlighted by the arrows in Figure 8(b)). This is in agreement with the conclusion discussed in Section 3.2, i.e., that the W2-norm EWI has the capability to solve hidden-layer problems.

Figures 9(a) and 9(b) display the corresponding waveform fitting between the observed data (black) and the synthetic data (red) associated with the L2-norm EWI and the W2-norm EWI of the common shot gather (CGS) at shot 50, respectively. The waveform data are significantly better matched by the W2-norm EWI than the L2-norm EWI. Figure 10 shows the normalized waveform misfit associated with L2-norm EWI and W2-norm EWI results at each stage during inversion, and the value of relative misfit function is shown in Table 1. It indicates that the W2-norm EWI significantly speeds up the inversion and converges faster than the L2-norm EWI at any stage.

5. Discussion

The method proposed in this work was designed to partly solve the problems relevant to L2-norm used as the objective function in the framework of EWI. The quadratic Wasserstein distance was used instead. The change in the misfit measurement between the predicted and observed data seismograms appears to produce some advantages, e.g., it is capable of mitigating cycle-skipping problems, allowing accurate convergence to the global minimum in the EWI. In addition, it was demonstrated to be insensitive to noise. These features indicate that the W2-norm EWI may be a more robust strategy for near-surface imaging than L2-norm EWI.

As is well known, nonuniqueness has been a fundamental and common issue in geophysical inversion, especially in EWI. Our work does not completely solve this problem, but it provides a good breakthrough point to study EWI solutions using the W2-norm for near-surface structures. Based on this framework, more waveform databases (e.g., first-arrival travel time and early arrival envelope) can be used for EWI and can be jointly inverted to produce improved near-surface solutions.

Wavelet extraction is critical in real data applications of FWI. Previous studies have reported that there are several methods to extract wavelets. Yilmaz [41] proposed that wavelet extraction can be conducted through deconvolution if the zero phase (or minimum phase) of the source wavelet is true. However, in reality, the source wavelet consists of mixed phase. Pratt [10] suggested that the source wavelet can be inverted during the first iteration if assuming that current model is accurate. Another way is to simultaneously invert the source wavelet and the velocity parameters in the FWI [42]. In this study, we stacked the first-arrival along the first breaks to extract the source wavelet from the real data [42].

Using the W2-norm in EWI benefits from the relatively simple and stable waveforms of early arrivals (arriving after the first arrival only within a few periods). In contrast, FWI using full waveform information has more challenges in real data applications because of the large number of practical problems, such as the waveform data-signal processing problems in signal-to-noise ratio improvement, the effect of rugged topography beneath real irregular receiver acquisition, a very complicated near-surface effect. Ignoring any issue in real data waveform inversion applications could cause the inversion results to be unexplained or unstable in FWI, which should be investigated further in future work.

6. Conclusions

In this study, we developed an early arrival waveform inversion based on the quadratic-Wasserstein-metric for near-surface imaging. First, we described the excellent properties of the Wasserstein metric that preserves the geometric characteristics of the probability distributions in detail. This feature results in the objective function based on the Wasserstein metric having a better convexity and helps mitigating the cycle-skipping problem in conventional FWI based on the L2-norm misfit function. Then, we gave the continuous form of the adjoint source and the Fréchet gradient under the W2-norm misfit function for seismic early arrival. The synthetic tests revealed that the method can mitigate cycle-skipping issues and map accurate near-surface velocity structure with high-resolution. In addition, the W2-norm EWI is capable of handling noise. The application on a marine dataset revealed that the solution of our new method can recover more details in the near-surface region, especially when dealing with hidden layers.

Nevertheless, the application of the W2-norm EWI is not sufficient to solve the nonuniqueness in geophysical inverse problems. More seismic data (e.g., travel time and envelopes) should be used in EWI to jointly invert an improved near-surface solution, which could provide a good initial model for FWI. In addition, topography variation in near-surface may produce artifacts leading to incorrect result during inversion; thus, the influence of rugged topography on EWI should be addressed and studied further in future work.

Data Availability

The data and the code in this paper can be obtained by contacting the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have influenced the work reported in this paper.

Acknowledgments

Supplementary Materials

There is a synthetic test on insensitivity to noise of the W2-norm EWI compared with L2-norm EWI, which is shown in Figure S1. (*Supplementary Materials*)

References

[1] D. Marsden, "Static corrections-a review, part III," *The Leading Edge*, vol. 12, no. 3, pp. 210–216, 1993.

[2] K. Aki and P. G. Richards, *Quantitative Seismology*, University Science Books, Sausalito, CA, USA, 2nd edition, 2002.

[3] R. G. Pratt and M. H. Worthington, "Inverse theory applied to multi-source cross-hole tomography part 1: acoustic wave-equation method1," *Geophysical Prospecting*, vol. 38, no. 3, pp. 287–310, 1990.

[4] X. Zhu, D. P. Sixta, and B. G. Angstman, "Tomostatics: t," *The Leading Edge*, vol. 11, no. 12, pp. 15–23, 1992.

[5] V. Cervený and J. E. P. Soares, "Fresnel volume ray tracing," *Geophysics*, vol. 57, no. 7, pp. 902–915, 1992.

[6] D. W. Vasco, J. E. Peterson, and E. L. Majer, "Beyond ray tomography: wavepaths and Fresnel volumes," *Geophysics*, vol. 60, no. 6, pp. 1790–1804, 1995.

[7] Y. Luo and G. T. Schuster, "Wave-equation traveltime inversion," *Geophysics*, vol. 56, no. 5, pp. 645–653, 1991.

[8] A. Tarantola, "Inversion of seismic reflection data in the acoustic approximation," *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984.

[9] A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Parameter Estimation*, SIAM, Pennsylvania, PA, USA, 1987.

[10] R. G. Pratt, Z.-M. Song, P. Williamson, and M. Warner, "Two-dimensional velocity models from wide-angle seismic data by wavefield inversion," *Geophysical Journal International*, vol. 124, no. 2, pp. 323–340, 1996.

[11] R. G. Gerhard Pratt, C. Shin, and G. J. Hicks, "Gauss-Newton and full Newton methods in frequency-space seismic wave-form inversion," *Geophysical Journal International*, vol. 133, no. 2, pp. 341–362, 1998.

[12] G. Yao, N. V. Da Silva, M. Warner, D. Wu, and C. Yang, "Tackling cycle skipping in full-waveform inversion with intermediate data," *Geophysics*, vol. 84, no. 3, pp. R411–R427, 2019.

[13] X. Li, G. Yao, F. Niu, D. Wu, and N. Liu, "Waveform inversion of seismic first arrivals acquired on irregular surface," *Geophysics*, vol. 87, no. 3, pp. R291–R304, 2022.

[14] J. Sheng, A. Leeds, M. Buddensiek, and G. Schuster, "Early arrival waveform tomography on near-surface refraction data," *Geophysics*, vol. 71, no. 4, pp. 47–57, 2006.

[15] C. Boonyasiriwat, G. T. Schuster, P. Valasek, and W. Cao, "Applications of multiscale waveform inversion to marine data using a flooding tech-nique and dynamic early-arrival windows," *Geophysics*, vol. 75, no. 6, pp. 129–136, 2010.

[16] Q. F. Xue, S. M. Hanafy, Y. K. Zheng, and Y. Wang, "Application of early arrival waveform inversion to Qademah shallow land data," *EAGE Technical Program Expanded Abstracts*, 2016.

[17] X. Shen, "Near-surface velocity estimation by weighted early-arrival waveform inversion," *SEG Technical Program Expanded Abstracts*, pp. 1975–1979, 2010.

[18] H. Yu, D. Zhang, and X. Wang, "Application of weighted early-arrival waveform inversion to shallow land data," *Journal of Applied Geophysics*, vol. 102, pp. 68–76, 2014.

[19] C. Zhou, W. Cai, Y. Luo, G. T. Schuster, and S. Hassanzadeh, "Acoustic wave-equation traveltime and waveform inversion of crosshole seismic data," *Geophysics*, vol. 60, no. 3, pp. 765–773, 1995.

[20] W. Zhou, R. Brossier, J. Virieux, and S. Operto, "Joint full-waveform inver-sion of early arrivals and reflections: a real OBC case study with gas cloud," in *Proceedings of the 86th Annual International Meeting, SEG Technical Program Expanded Abstracts*, pp. 1247–1251, Dallas, TX, USA, September 2016.

[21] Y. Ma and D. Hale, "Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion," *Geophysics*, vol. 78, no. 6, pp. R223–R233, 2013.

[22] R. S. Wu, J. Luo, and B. Wu, "Ultra-low-frequency information in seismic data and envelope inversion," *SEG Technical Program Expanded Abstracts*, pp. 3078–3082, 2013.

[23] E. Bozdag, J. Trampert, and J. Tromp, "Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements," *Geophysical Journal International*, vol. 185, pp. 845–870, 2014.

[24] K. Jiao, D. Sun, X. Cheng, and D. Vigh, "Adjustive full waveform inversion. 85th annual international meeting," *SEG Technical Program Expanded Abstracts*, pp. 1091–1095, 2015.

[25] M. Warner and L. Guasch, "Adaptive waveform inversion: theory," *SEG Technical Program Expanded Abstracts 2014*, vol. 2014, pp. 1–5, 2014.

[26] C. Zhang, H. Yao, Q. Liu et al., "Linear array ambient noise adjoint tomography reveals intense crust-mantle interactions in north China c," *Journal of Geophysical Research: Solid Earth*, vol. 123, no. 1, pp. 368–383, 2018.

[27] Y. N. Yang, B. Engquist, J. Z. Sun, and B. F. Hamfeldt, "Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion," *Geophysics*, vol. 83, no. 1, pp. 43–62, 2018.

[28] B. Engquist, B. D. Froese, and Y. Yang, "Optimal transport for seismic full waveform inversion," *Communications in Mathematical Sciences*, vol. 14, no. 8, pp. 2309–2330, 2016.

[29] L. Métivier, R. Brossier, Q. Mrigot, E. Oudet, and J. Virieux, "Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion," *Geophysical Journal International*, vol. 205, pp. 345–377, 2016.

[30] B. Engquist and B. D. Froese, "Application of the Wasserstein metric to seismic signals," *Communications in Mathematical Sciences*, vol. 12, no. 5, pp. 979–988, 2014.

[31] C. Villani, *Topics in Optimal Transportation. Graduate Studies in Mathematics*, American Mathematical Society, Providence, RH, USA, 2003.

[32] B. Sun and T. Alkhalifah, "The application of an optimal transport to a preconditioned data matching function for robust waveform inversion," *Geophysics*, vol. 84, no. 6, pp. R923–R945, 2019.

[33] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Communications on Pure and Applied Mathematics*, vol. 44, no. 4, pp. 375–417, 1991.

[34] G. Monge, *Sur le calcul intégral des équations aux differences partielles*, Mémoires del'Académie des Sciences, London, UK, 1784.

[35] A. M. Ampere, "Mémoire contenant l'application de la théorie," *Journal de l'École Poly technique*, 1820.

[36] X. Gu, F. Luo, J. Sun, and S.-T. Yau, "Variational principles for Minkowski type problems, discrete optimal transport, and discrete Monge-Ampère equations," *Asian Journal of Mathematics*, vol. 20, no. 2, pp. 383–398, 2016.

[37] K. Soheil, S. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, "Optimal mass transport signal processing and machine-learning applications," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5258–5267, Honolulu, HI, USA, July 2017.

[38] A. Maggi, C. Tape, M. Chen, D. Chao, and J. Tromp, "An automated time-window selection algorithm for seismic tomography," *Geophysical Journal International*, vol. 178, no. 1, pp. 257–281, 2009.

[39] J. Nocedal, "Updating quasi-Newton matrices with limited storage," *Mathematics of Computation*, vol. 35, no. 151, pp. 773–782, 1980.

[40] R. Schmoller, "Some aspects of handling velocity inversion and hidden layer problems in seismic refraction work," *Geophysical Prospecting*, vol. 30, no. 6, pp. 735–751, 1982.

[41] O. Yilmaz, *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seis-Mic Data*, SEG, Tulsa, Oklahoma, 2001.

[42] J. Rickett, "The variable projection method for waveform inversion with an unknown source function," *Geophysical Prospecting*, vol. 61, no. 4, pp. 874–881, 2013.