Numer. Math. Theor. Meth. Appl.
doi: 10.4208/nmtma.OA-2022-0111

Vol. **16**, No. 2, pp. 277-297
May 2023

On Quadratic Wasserstein Metric with Squaring Scaling for Seismic Velocity Inversion

Zhengyang Li[1], Yijia Tang[2], Jing Chen[3,*] and Hao Wu[1]

[1] *Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China*
[2] *School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China*
[3] *Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 639798*

Received 14 July 2022; Accepted (in revised version) 30 January 2023

Abstract. The quadratic Wasserstein metric has shown its power in comparing probability densities. It is successfully applied in waveform inversion by generating objective functions robust to cycle skipping and insensitive to data noise. As an alternative approach that converts seismic signals to probability densities, the squaring scaling method has good convexity and thus is worth exploring. In this work, we apply the quadratic Wasserstein metric with squaring scaling to regional seismic tomography. However, there may be interference between different seismic phases in a broad time window. The squaring scaling distorts the signal by magnifying the unbalance of the mass of different seismic phases and also breaks the linear superposition property. As a result, illegal mass transportation between different seismic phases will occur when comparing signals using the quadratic Wasserstein metric. Furthermore, it gives inaccurate Fréchet derivative, which in turn affects the inversion results. By combining the prior seismic knowledge of clear seismic phase separation and carefully designing the normalization method, we overcome the above problems. Therefore, we develop a robust and efficient inversion method based on optimal transport theory to reveal subsurface velocity structures. Several numerical experiments are conducted to verify our method.

AMS subject classifications: 49N45, 65K10, 86-08, 86A15

Key words: Optimal transport, Wasserstein metric, waveform inversion, seismic velocity inversion, squaring scaling.

*Corresponding author. *Email addresses:* hwu@tsinghua.edu.cn (H. Wu), yijia_tang@sjtu.edu.cn (Y.J. Tang), jing.chen@ntu.edu.sg (J. Chen), lizhengy18@mails.tsinghua.edu.cn (Z.Y. Li)

1. Introduction

Seismic waveform inversion has been receiving wide attention in past decades [4, 14, 29, 32, 38, 42, 46, 47] due to its high-resolution imaging capability. The major goal is to find optimal model parameters that minimize the discrepancy between synthetic and observed seismic signals. In mathematics, it can be formulated as a partial differential equation (PDE) constrained optimization problem, which consists of two key ingredients [41]: the forward modeling of wave propagation and updating model parameters. In previous decades, limited by computing power, most tomography methods simulated wave propagation based on the ray theory. The high frequency assumption ignores finite frequency phenomena such as wave-front healing and scattering [17], and thus, leads to low-resolution inversion results [32]. With the rapid development of computing power and the forward modeling method, more accurate synthetic signals could be computed by numerically solving wave equations. It enables us to obtain high-resolution subsurface velocity structures from the inversion of waveform data, which could provide guidance information for seismic hazard assessment [37] and exploration geophysics [41].

The discrepancy between synthetic and observed seismic waveforms is usually measured using the L^2 metric [32, 37, 38, 41]. However, it suffers from the well-known cycle skipping problem [41] so that the solution may be trapped in local minima during the iteration, leading to incorrect inversion results. To overcome the problem, many methods have been proposed to modify the objective function, e.g., the envelope objective function [5], the cross-correlation-based objective function [23], and the deconvolution-based objective function [22]. In addition, the quadratic Wasserstein (W_2) metric from the Optimal Transport (OT) theory [36, 39, 40] has received wide attention in recent years due to its nice properties and has been applied to many seismic inverse problems such as earthquake location and seismic tomography [7, 10–13, 46, 47, 49]. This metric measures the difference between two probability distributions by minimizing the transport cost from one distribution to the other, which is insensitive to data noise and preserves the convexity regarding data shift, dilation, and partial amplitude change [10, 11, 14]. Thanks to these advantages, the reconstruction of velocity models can succeed even if the initial model is far from the real model [14, 27, 28].

However, it is not straightforward to apply the quadratic Wasserstein metric in the seismic waveform inversion [13]. The main reason is that the seismic signals are signed. Thus, a key ingredient in the application of the quadratic Wasserstein metric to seismic waveform inversion is converting seismic waveforms to probability distributions. Various scaling techniques are developed to deal with this problem, e.g., linear scaling [47], squaring scaling [7], exponential scaling [33], and graph-space transform [26]. Moreover, there are also some other metrics based on the OT theory that have been applied to seismic inverse problems, e.g., the Wasserstein-Fisher-Rao metric and the Kantorovich-Rubinstein norm, which relax the mass conservation constraint [27, 28, 49]. Though the quadratic Wasserstein metric has been applied widely

in the exploration-scale inversion such as the Marmousi model [13, 28, 47] and the BP 2004 model [27, 46, 47], its application in regional-scale or global-scale waveform inversion is still under exploration [15].

In this work, we would like to provide a comprehensive study on the velocity inversion in regional-scale models based on squaring scaling and the quadratic Wasserstein metric. Theoretically, the quadratic Wasserstein metric with squaring scaling can generate an objective function with a smooth Fréchet derivative with respect to signal and incorporate well with the adjoint state methods [13]. Compared with other scaling methods, it has perfect convexity properties with respect to time shifts [46]. Due to these benefits, the property and the application of the quadratic Wasserstein metric with squaring scaling becomes a topic of great interest, which has been extensively studied in recent years [12–14, 46, 47].

However, this approach has been less successful than expected in velocity inversion at both exploration and regional scales. The signal in a broad time window usually contains multiple seismic phases, between which may exist interference. The squaring scaling magnifies the unbalance of the mass of different seismic phases and also breaks the linear superposition property. As a result, the quadratic Wasserstein metric suffers from interference between different seismic phases. The illegal mass transportation between different seismic phases occurs, resulting in the contamination of the sensitivity kernel and further leading to inaccurate inversion results. In exploration scale, the subsurface structure is usually complex [48], inducing multiple seismic phases that can not be separated. It is difficult to avoid interference between different seismic phases, which may be the reason why the quadratic Wasserstein metric with squaring scaling has not been successful in the exploration field. Fortunately, in regional velocity inversion, the main seismic phases may be separated [8, 16, 25] because the subsurface structure in regional scale is not as complex as that in exploration scale. This gives us the possibility to use the quadratic Wasserstein metric with squaring scaling to reveal subsurface velocity structures in regional scale problems.

A consensus discussion [12] argues that the failure of the quadratic Wasserstein metric with squaring scaling attributes to the squaring scaling that destroys the uniqueness since $f^2 = g^2$ does not imply $f = g$. However, in our previous study of the earthquake location [7], this non-uniqueness issue is partially avoided using the quadratic Wasserstein metric with squaring scaling, which is also recognized by Engquist and Yang [13]. Starting with an appropriate initial model, the generated synthetic signal g will be more similar to the real signal f than its opposite $-f$. By considering the prior seismic phase information and the function continuity constraint, the synthetic signal g tends to converge to the real signal f, ensuring the correct convergence. This success may show that the non-uniqueness issue of squaring scaling is not the most dominant reason for the unsuccessful inversion. In fact, the non-uniqueness issue widely exists in the velocity inversion problems but does not lead to the failure of inversion, e.g., [2, 23], in which the misfit equals 0 does not ensure the consistency of real and synthetic signals. Despite the non-uniqueness issue, there are still more fundamental issues that need to be investigated. According to our later observation in Section 3,

squaring scaling magnifies the unbalance of the mass of different signal components. This leads to illegal mass transportation between different seismic phases, which rarely occurs even for the classical L^2 metric [13]. As a result, the nice convexity property of the objective function [11] cannot be maintained. Thus, the Fréchet derivative obtained by the above process is also inaccurate or wrong. This also affects the inversion process and results. Moreover, we find that the rough normalization technique can also lead to illegal mass transportation. Worse yet, it can even create false seismic phases. These are all problems we need to overcome.

Based on the above understandings, we exploit and develop techniques for the velocity inversion problem using the quadratic Wasserstein metric with modified squaring scaling. Our efforts revolve around avoiding illegal mass transportation. Specifically, we combine the prior knowledge of seismology to match the same seismic phases. This is a common technique in seismology. But it could work very well with the quadratic Wasserstein metric rather than the classical L^2 metric, especially for the well-known cycle-skipping phenomenon. This improves the robustness of the inversion method. Another study in this manuscript is that we carefully design the normalization method. For the quadratic Wasserstein metric, the appropriate normalization method is very important. We not only need to avoid the singularity in the Fréchet derivative produced by the inappropriate optimal transport map but also ensure that the false seismic phases and illegal mass transportation do not appear. Based on the above strategies, we propose an effective velocity inversion method by using the quadratic Wasserstein metric with modified squaring scaling, which is capable of revealing the velocity structure accurately and efficiently in regional-scale tomography. We have to admit that there are also other scaling methods that work well with the quadratic Wasserstein metric for velocity inversion. However, the squaring scaling method is considered to have good convexity. We would like to explore it in-depth and look forward to its advantages in future research. This is the motivation for our work.

The rest of the paper is organized as follows. In Section 2, we briefly review the mathematical formula of seismic velocity inversion and the basics of the quadratic Wasserstein metric. We discuss important issues in the inversion and present detailed implementations in Section 3. Meanwhile, we illustrate the necessity of our method by some toy models. In Section 4, the numerical experiments are conducted to demonstrate the effectiveness and efficiency of our method. Finally, we conclude the paper in Section 5.

2. The quadratic Wasserstein metric and seismic velocity inversion

We review the seismic waveform tomography and the adjoint state method in this section. The seismic velocity inversion can be written as the PDE constrained optimization problem

$$c_T(\boldsymbol{x}) = \arg\min_{c(\boldsymbol{x})} \Xi\big(c(\boldsymbol{x})\big), \quad \Xi\big(c(\boldsymbol{x})\big) = \sum_{i=1}^{N}\sum_{j=1}^{M} \chi_{ij}\big(c(\boldsymbol{x})\big), \tag{2.1}$$

where index (i,j) indicates the source-receiver pair. We consider N earthquakes recorded by M seismic stations. Correspondingly, the misfit function χ_{ij} is defined as

$$\chi_{ij}\big(c(\boldsymbol{x})\big) = \mathcal{D}\big(s_{ij}\big(t;c(\boldsymbol{x})\big), d_{ij}(t)\big). \tag{2.2}$$

Here, $\mathcal{D}$ is the distance function that measures the difference between the real seismic signal $d_{ij}(t)$ and the synthetic signal $s_{ij}(t;c(\boldsymbol{x}))$. These two signals can be regarded as the wavefields at the station

$$d_{ij}(t) = u_i\big(\boldsymbol{\eta}_j, t; c_T(\boldsymbol{x})\big), \quad s_{ij}\big(t;c(\boldsymbol{x})\big) = u_i\big(\boldsymbol{\eta}_j, t; c(\boldsymbol{x})\big) \tag{2.3}$$

satisfying the following acoustic wave equation with initial boundary conditions:

$$\begin{aligned}\frac{\partial^2 u_i\big(\boldsymbol{x},t;c(\boldsymbol{x})\big)}{\partial t^2} &= \nabla\cdot\big(c^2(\boldsymbol{x})\nabla u_i\big(\boldsymbol{x},t;c(\boldsymbol{x})\big)\big)\\ &\quad + R(t-\tau_i)\delta(\boldsymbol{x}-\boldsymbol{\xi}_i), \qquad \boldsymbol{x}\in\Omega, \quad t>0, \end{aligned}\tag{2.4}$$

$$u_i\big(\boldsymbol{x},0;c(\boldsymbol{x})\big) = \frac{\partial u_i\big(\boldsymbol{x},0;c(\boldsymbol{x})\big)}{\partial t} = 0, \qquad \boldsymbol{x}\in\Omega, \tag{2.5}$$

$$\boldsymbol{n}\cdot\big(c^2(\boldsymbol{x})\nabla u_i\big(\boldsymbol{x},t;c(\boldsymbol{x})\big)\big) = 0, \qquad \boldsymbol{x}\in\partial\Omega, \quad t>0. \tag{2.6}$$

Here, the locations of the earthquake and receiver station are $\boldsymbol{\xi}_i$ and $\boldsymbol{\eta}_j$, the origin time of the earthquake is τ_i. The seismic rupture is modeled as the point source $\delta(\boldsymbol{x}-\boldsymbol{\xi}_i)$ by assuming its scale is much smaller compared to the scale of seismic wave propagation [1,24]. And the source time function is simplified as the Ricker wavelet

$$R(t) = A\big(1-2\pi^2 f_0^2 t^2\big)e^{-\pi^2 f_0^2 t^2}, \tag{2.7}$$

where f_0 denotes the dominant frequency, and A is the normalization factor. The outward unit normal vector to the simulation domain boundary $\partial\Omega$ is $\boldsymbol{n}$. In practice, the perfectly matched layer absorbing boundary condition [20] is used to deal with the propagation of waves outside the area. In this section, we use the reflection boundary condition to simplify the derivation.

Remark 2.1. Here, we consider the trace by trace strategy to apply the 1D quadratic Wasserstein metric to the waveform inversion. This is an effective approach for geological scale inversion [7,47], and also much easier in mathematics since the 1D quadratic Wasserstein metric admits a closed form expression [36,39,40].

2.1. The adjoint method

Below, we briefly review the adjoint method [11,31] for solving the optimization problems (2.1)-(2.7). We assume that the small perturbation δc of seismic velocity structure causes the perturbation of the wavefield

$$\delta u_i\big(\boldsymbol{x},t;c(\boldsymbol{x})\big) = u_i\big(\boldsymbol{x},t;c(\boldsymbol{x})+\delta c(\boldsymbol{x})\big) - u_i\big(\boldsymbol{x},t;c(\boldsymbol{x})\big). \tag{2.8}$$

For the sake of brevity, we will omit the parameter $c(\boldsymbol{x})$ of the wavefield and the signals in the following. The perturbation $\delta u_i(\boldsymbol{x},t)$ satisfies the equations

$$\frac{\partial^2 \delta u_i(\boldsymbol{x},t)}{\partial t^2} = \nabla\cdot\big(c^2(\boldsymbol{x})\nabla\delta u_i(\boldsymbol{x},t)\big) + \nabla\cdot\big(\big(2c(\boldsymbol{x})+\delta c(\boldsymbol{x})\big)\delta c(\boldsymbol{x})\nabla(u_i+\delta u_i)(\boldsymbol{x},t)\big), \qquad \boldsymbol{x}\in\Omega, \tag{2.9}$$

$$\delta u_i(\boldsymbol{x},0) = \frac{\partial \delta u_i(\boldsymbol{x},0)}{\partial t} = 0, \qquad \boldsymbol{x}\in\Omega, \tag{2.10}$$

$$\boldsymbol{n}\cdot\big(c^2(\boldsymbol{x})\nabla\delta u_i(\boldsymbol{x},t) + \big(2c(\boldsymbol{x})+\delta c(\boldsymbol{x})\big)\delta c(\boldsymbol{x})\nabla(u_i+\delta u_i)(\boldsymbol{x},t)\big) = 0, \quad \boldsymbol{x}\in\partial\Omega. \tag{2.11}$$

Multiply the test function $w_i(\boldsymbol{x},t)$ on Eq. (2.9) and integrate it on $\Omega\times[0,t_f]$ for sufficient large time t_f. Using integration by parts yields

$$\begin{aligned}
&\int_0^{t_f}\int_\Omega \frac{\partial^2 w_i}{\partial t^2}\delta u_i \mathrm{d}\boldsymbol{x}\mathrm{d}t - \int_\Omega \frac{\partial w_i}{\partial t}\delta u_i\Big|_{t=t_f}\mathrm{d}\boldsymbol{x} + \int_\Omega w_i\frac{\partial \delta u_i}{\partial t}\Big|_{t=t_f}\mathrm{d}\boldsymbol{x}\\
&= \int_0^{t_f}\int_\Omega \nabla\cdot(c^2\nabla w_i)\delta u_i \mathrm{d}\boldsymbol{x}\mathrm{d}t - \int_0^{t_f}\int_{\partial\Omega}\boldsymbol{n}\cdot(c^2\nabla w_i)\delta u_i\mathrm{d}\zeta\mathrm{d}t\\
&\quad - \int_0^{t_f}\int_\Omega (2c+\delta c)\,\delta c\nabla w_i\cdot\nabla(u_i+\delta u_i)\mathrm{d}\boldsymbol{x}\mathrm{d}t\\
&\approx \int_0^{t_f}\int_\Omega \nabla\cdot(c^2\nabla w_i)\delta u_i \mathrm{d}\boldsymbol{x}\mathrm{d}t - \int_0^{t_f}\int_{\partial\Omega}\boldsymbol{n}\cdot(c^2\nabla w_i)\delta u_i\mathrm{d}\zeta\mathrm{d}t\\
&\quad - \int_0^{t_f}\int_\Omega 2c\delta c\nabla w_i\cdot\nabla u_i\mathrm{d}\boldsymbol{x}\mathrm{d}t,
\end{aligned} \tag{2.12}$$

where the higher-order terms are ignored in the last step since we can naturally assume that $\|\delta u_i\| \ll \|u_i\|$ and $\|\delta c(\boldsymbol{x})\| \ll \|c(\boldsymbol{x})\|$.

On the one hand, the perturbation of misfit $\delta\chi_{ij}$ results from the wave speed perturbation $\delta c(\boldsymbol{x})$, which writes

$$\begin{aligned}
\delta\chi_{ij}(c) &= \mathcal{D}\big(s_{ij}(t)+\delta s_{ij}(t), d_{ij}(t)\big) - \mathcal{D}\big(s_{ij}(t), d_{ij}(t)\big)\\
&\approx \big\langle Q_{ij}(t), \delta s_{ij}(t)\big\rangle = \int_0^{t_f} Q_{ij}(t)\delta s_{ij}(t)\mathrm{d}t.
\end{aligned}$$

Here, $Q_{ij}(t)$ indicates the Fréchet derivative of the distance $\mathcal{D}$ with respect to the synthetic data $s_{ij}(t)$

$$Q_{ij}(t) = \nabla_s\mathcal{D}(s,d)\big|_{s=s_{ij}(t),d=d_{ij}(t)}, \tag{2.13}$$

which will be specified later. Let $w_i(\boldsymbol{x},t)$ satisfy the adjoint equation

$$\frac{\partial^2 w_i(\boldsymbol{x},t)}{\partial t^2} = \nabla\cdot\big(c^2(\boldsymbol{x})\nabla w_i(\boldsymbol{x},t)\big) + \sum_{j=1}^{M} Q_{ij}(t)\delta(\boldsymbol{x}-\boldsymbol{\eta}_j), \quad \boldsymbol{x}\in\Omega, \tag{2.14}$$

$$w_i(\boldsymbol{x},t_f) = \frac{\partial w_i(\boldsymbol{x},t_f)}{\partial t} = 0, \qquad \boldsymbol{x}\in\Omega, \tag{2.15}$$

$$\boldsymbol{n}\cdot\big(c^2(\boldsymbol{x})\nabla w_i(\boldsymbol{x},t)\big) = 0, \qquad \boldsymbol{x}\in\partial\Omega. \tag{2.16}$$

Multiply $\delta u_i(\boldsymbol{x},t)$ on Eq. (2.14), integrate it on $\Omega\times[0,t_f]$ and subtract (2.12) to obtain

$$\begin{aligned}\sum_{j=1}^{M}\int_0^{t_f} Q_{ij}(t)\delta s_{ij}(t)\mathrm{d}t &= \sum_{j=1}^{M}\int_0^{t_f}\int_\Omega Q_{ij}(t)\delta(\boldsymbol{x}-\boldsymbol{\eta}_j)\delta u_i(\boldsymbol{x},t)\mathrm{d}\boldsymbol{x}\mathrm{d}t\\ &= -\int_0^{t_f}\int_\Omega 2c(\boldsymbol{x})\delta c(\boldsymbol{x})\nabla w_i(\boldsymbol{x},t)\cdot\nabla u_i(\boldsymbol{x},t)\mathrm{d}\boldsymbol{x}\mathrm{d}t.\end{aligned}$$

The linear relationship between $\delta\Xi$ and $\delta c(\boldsymbol{x})$ is established as

$$\delta\Xi(c)=\sum_{i=1}^{N}\sum_{j=1}^{M}\delta\chi_{ij}(c)=\sum_{i=1}^{N}\int_\Omega K_i(\boldsymbol{x})\delta c(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \tag{2.17}$$

where the sensitivity kernel of the i-th source for $c(\boldsymbol{x})$ is defined as

$$K_i(\boldsymbol{x})=-\int_0^{t_f}2c(\boldsymbol{x})\nabla w_i(\boldsymbol{x},t)\cdot\nabla u_i(\boldsymbol{x},t)\mathrm{d}t. \tag{2.18}$$

2.2. The quadratic Wasserstein metric

As we discussed at the beginning of this section, the synthetic signal $s_{ij}(t)$ and real seismic signal $d_{ij}(t)$ are time series. As we know, the quadratic Wasserstein metric between the 1D probability density functions has an analytic form [7,39,40,47]

$$W_2^2(f,g)=\int_0^{t_f}|t-T(t)|^2 f(t)\mathrm{d}t,\quad T(t)=G^{-1}\big(F(t)\big). \tag{2.19}$$

Here $f(t),g(t)$ are probability density functions defined on $[0,t_f]$ and $F(t),G(t)$ are corresponding cumulative density functions

$$F(t)=\int_0^t f(\tau)\mathrm{d}\tau,\quad G(t)=\int_0^t g(\tau)\mathrm{d}\tau.$$

Note that the seismic signals are not probability density functions. We need to transform them into nonnegative and normalized functions for the quadratic Wasserstein metric comparison. In other words, the misfit function defined in (2.2) can be written as

$$\chi_{ij}=\mathcal{D}\big(s_{ij}(t),d_{ij}(t)\big)=W_2^2\big(\mathcal{P}\big(s_{ij}(t)\big),\mathcal{P}\big(d_{ij}(t)\big)\big). \tag{2.20}$$

The operator $\mathcal{P}$ converts the seismic signals into probability density functions, which will be discussed in Section 3. Thus, we can obtain the expression of the Fréchet derivative [7,47] mentioned in (2.13)

$$\begin{aligned}\nabla_s\mathcal{D}(s,d)&=\nabla_f W_2^2(f,g)\big|_{f=\mathcal{P}(s),g=\mathcal{P}(d)}\cdot\nabla_s\mathcal{P}(s)\\ &=\left\langle 2\int_0^t\tau-T(\tau)\mathrm{d}\tau,\nabla_s\mathcal{P}(s)\right\rangle.\end{aligned} \tag{2.21}$$

3. Data preprocessing and new normalization

One key step in solving the optimization problem is to calculate accurate sensitivity kernels for updating model parameters. As shown in Eq. (2.21), it highly depends on the calculation of the optimal transport map $T(\tau)$. Thus, it is worth discussing how to construct a reasonable and accurate optimal transport map.

Firstly, a seismic signal usually contains several seismic phases. When comparing the whole signals using the quadratic Wasserstein metric with squaring scaling, the unbalanced mass of different phases is magnified, which will lead to mass transportation between different seismic phases. This illegal mass transportation will generate an unreasonable optimal transport map. Accordingly, we limit the mass transportation within the same phase by applying time windows. This strategy focuses on the signal comparison between the same phase, which derives appropriate optimal transport maps.

Secondly, to avoid the numerical singularities in the computation of the transport map (2.19), it is required to assign a small constant to the squared signals to ensure they are strictly positive and make the corresponding cumulative distribution functions strictly monotone (see [36, Remark 2.6]). However, rough normalization techniques, such as adding the same constant and then normalizing, would create false seismic phases because different masses are introduced to the compared signals. As a result, the illegal mass transportation will occur between the signal and the assigned mass, leading to unreasonable optimal transport maps. To address this issue, we carefully design the normalization method that not only avoids the singularity but also eliminates the false seismic phases and illegal mass transportation.

To have a better understanding of the two issues, we illustrate the necessity of our method by the classical two-layer model [14] modified from the global reference models [18, 19]. The initial and real seismic velocity models are shown in Fig. 1. The real model is bounded in $\Omega = [0, 80\ \mathrm{km}] \times [0, 60\ \mathrm{km}]$, consisting of the crust and the uppermost mantle that separated by the Moho discontinuity at $30\ \mathrm{km}$ depth. A cubic $+15\%$ high velocity anomaly is set in the crust, given by

$$c_T(x,z) = \begin{cases} 6.67\ \mathrm{km/s}, & (x,z) \in [35\ \mathrm{km}, 45\ \mathrm{km}] \times [10\ \mathrm{km}, 20\ \mathrm{km}], \\ 8.1\ \mathrm{km/s}, & z > 30\ \mathrm{km}, \\ 5.8\ \mathrm{km/s}, & \text{otherwise}. \end{cases}$$

Correspondingly, the initial velocity model without high-velocity anomaly is as follows:

$$c_0(x,z) = \begin{cases} 5.8\ \mathrm{km/s}, & z \le 30\ \mathrm{km}, \\ 8.1\ \mathrm{km/s}, & z > 30\ \mathrm{km}. \end{cases}$$

Our goal is to perform the seismic velocity inversion to reveal the high-velocity anomaly.

The computational time interval is $[0\ \mathrm{s}, 21\ \mathrm{s}]$. The inversion grid spacing is $2\ \mathrm{km}$ and the number of degrees of freedom amounts to 1200. The space and time steps in

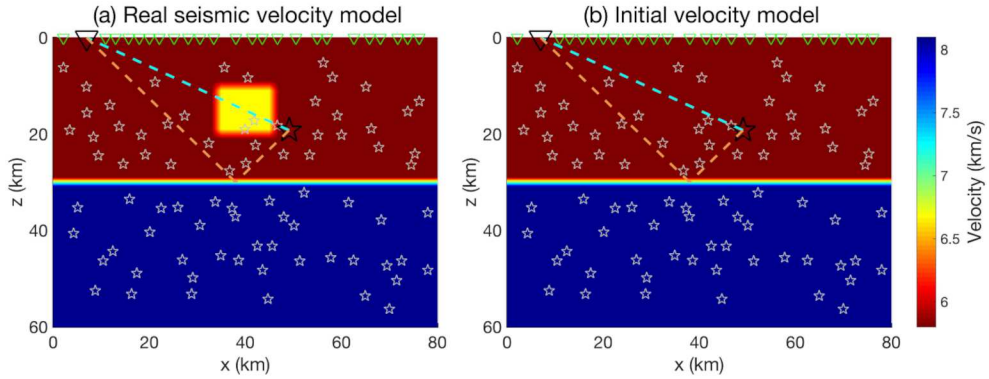

Figure 1: Illustration of the two-layer models. (a) The real seismic velocity model with a high-velocity anomaly. (b) The initial velocity model. The green inverted triangles indicate the receiver stations and the white stars indicate the earthquakes. The specific source-receiver pair is highlighted by the black star and inverted triangle. The cyan and tan dashed lines are the direct wave path and the reflected wave path, respectively.

the forward simulation are 0.2 km and 0.01 s, respectively. The dominant frequency of the source time function in (2.7) is $f_0 = 2$ Hz. We randomly choose 25 receiver stations deployed on the surface ($z = 0$ km) and 80 earthquakes distributed in the study region.

3.1. Selecting source-receiver pairs and picking seismic phases

Due to subsurface velocity discontinuities such as the Moho interface, a seismic signal may contain different seismic phases, including the direct wave and the reflected wave. These seismic waves propagate along different wave paths and carry distinct underground structure information. Sometimes, the direct wave and the reflected wave arrive simultaneously and can not be discriminated, which is known as the multipath phenomenon [34]. It is not trivial to extract robust information from this kind of constraint. One common technique in seismology is to manually exclude these source-receiver pairs to avoid interference caused by unreliable constraints [8, 16]. Here we also follow this strategy. When different seismic phases can be identified, the same phases of the real seismic signal and the synthetic signal should be matched when comparing these signals. However, the unreasonable match between different phases usually happens in the quadratic Wasserstein metric comparison due to the mass conservation and global matching property [13, 46]. When the masses of the real seismic signal and the synthetic signal are unbalanced for the same phase, illegal mass transportation between different seismic phases occurs. In this case, the quadratic Wasserstein metric will break the convexity of the objective function and produce an unreasonable optimal transport map, which will further result in artifacts in the sensitivity kernels. Particularly, the squaring scaling enhances the unbalance of the mass of different seismic phases, magnifying this problem [13]. Therefore, we need to deal with this issue seriously.

Many techniques in seismology have been proposed to ensure the match between the same phase, e.g., manually picking the signals and phases [8, 16], automated picking with the waveform simulations [25, 37], and deep-learning-based picking method [30, 35]. These methods pick the same seismic phases for comparison and exclude multi-path signals to avoid interference. In our study, we follow the same idea and use a simple approach to process the signals. We manually retain the signals whose seismic phases can be identified and select appropriate time windows that contain the same seismic phases according to theoretical arrival times [8, 43]. As a result, the same seismic phases of the real and synthetic signals are compared using the quadratic Wasserstein metric, which overcomes the cycle-skipping problem and generates reliable and reasonable optimal transport maps.

Next, we explain the necessity of the above-mentioned data preprocessing method. The squaring scaling operation $\mathcal{P}_2$ in Eq. (3.2) is applied for the computation of adjoint source and sensitivity kernel. For the earthquakes above the Moho discontinuity, at least two waves, the direct wave and the reflected wave can be recorded. Their wave paths are denoted by cyan dashed lines and tan dashed lines in Fig. 1, respectively. In the real seismic velocity model, the amplitude of the direct wave signal is slightly smaller since its energy dissipates when passing through the high-velocity anomaly. On the other hand, the reflected wave signal should be the same since the velocity structures around the reflected wave path are the same in the initial and real seismic velocity models (Fig. 1).

The difference between the real seismic signal and the synthetic signal is further magnified by the squaring scaling. It leads to illegal mass transportation from the direct wave of the synthetic signal to the reflected wave of the real seismic signal (Fig. 2(a)), causing the additional energy in the adjoint source in the interval of $11.5 - 13.0$ s in Fig. 2(e) and the artifacts in the sensitivity kernel shown in Fig. 2(b). If we only consider direct waves for inversion, the above-mentioned difficulties will be easily solved, as we illustrate in Figs. 2(c) and 2(d).

Remark 3.1. In practical application, the reflected wave signals are also important to constrain the underground velocity structures [16]. When the reflected waves are picked, our method can also be implemented using the reflected waves. However, the utilization of the reflected wave is not trivial, and more technical details are required in practice [3, 45, 50]. Thus, we will not discuss the usage of reflected waves in this work.

3.2. New normalization method

As it is well known, the quadratic Wasserstein metric measures the difference between two probability density functions, which is not directly suitable for seismic signals. Thus, some processing procedures, i.e., choosing an appropriate operator $\mathcal{P}$ in (2.20), are required to convert seismic signals into probability density functions. Several different approaches, e.g., linear scaling [47], squaring scaling [7], exponential

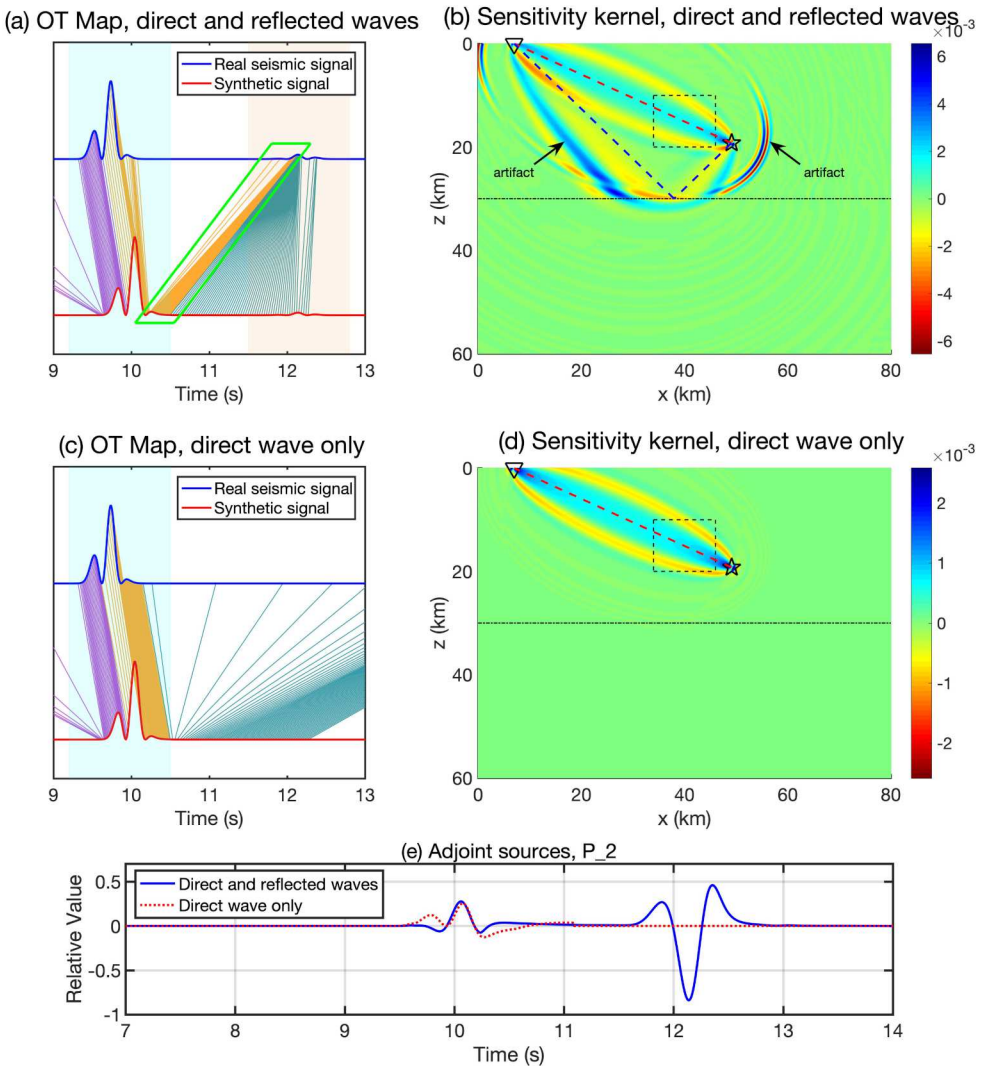

Figure 2: Illustration of the optimal transport maps between the real seismic signal and synthetic signal (a), (c), the sensitivity kernels (b), (d), and the comparison of adjoint sources (e). In (a) and (b), direct and reflected waves are compared simultaneously. In (c) and (d), only direct waves are compared. In (a) and (c), the blue and brown shaded areas denote the time window of the direct and reflected waves, respectively. The purple, yellow, and cyan lines represent the mass transportation from the synthetic signal at time t to the real seismic signal at time $T(t)$ according to the optimal transport plan in Eq. (2.19). The different colors correspond to the different time ranges. The illegal mass transportation from the direct wave of the synthetic signal to the reflected wave of the real seismic signal is within the green box in (a). In (b) and (d), the earthquake and the station receiver are represented by the black star and inverted triangle, respectively. The velocity anomaly is within the black dashed box. In (e), the adjoint sources of cases (a) and (c) based on the operator $\mathcal{P}_2$ in Eq. (3.2) are plotted for comparison.

scaling [33], and graph-space transform [26] have been proposed to address this issue. Among these methods, the squaring scaling is considered to have good convexity, which is worthy of in-depth discussions.

The normalization operator with squaring scaling consists of two ingredients: squaring seismic signal to ensure non-negativity and normalization to guarantee the same mass. A natural approach is

$$\mathcal{P}_1\big(s(t)\big) = \frac{s^2(t)}{\langle s^2(t)\rangle}, \tag{3.1}$$

in which

$$\langle s^2(t)\rangle = \int_0^{t_f} s^2(\tau)\mathrm{d}\tau.$$

Substitute the above formula into Eq. (2.20), the form of the misfit function is given by

$$\chi = \mathcal{D}\big(s(t), d(t)\big) = W_2^2\left(\frac{s^2(t)}{\langle s^2(t)\rangle}, \frac{d^2(t)}{\langle d^2(t)\rangle}\right).$$

Here the subscript indices i and j are dropped for simplicity. According to the discussions in Section 2.2, we need to compute the inverse of the following cumulative distribution function:

$$G(t) = \int_0^t \frac{d^2(\tau)}{\langle d^2(t)\rangle}\mathrm{d}\tau.$$

However, $G^{-1}(t)$ is not well defined when the real seismic signal $d(t) = 0$ in a certain interval. Accordingly, there will be difficulties in the computation of the optimal transport map and the misfit function.

In order to avoid the above-mentioned problem, we can make a slight upward shift on the squared signal before the normalization, i.e.,

$$\mathcal{P}_2\big(s(t)\big) = \frac{s^2(t)+\varepsilon}{\langle s^2(t)+\varepsilon\rangle}. \tag{3.2}$$

Here $\varepsilon > 0$ is a small parameter. However, the misfit function in (2.20) with this operator

$$\chi = \mathcal{D}\big(s(t), d(t)\big) = W_2^2\left(\frac{s^2(t)+\varepsilon}{\langle s^2(t)+\varepsilon\rangle}, \frac{d^2(t)+\varepsilon}{\langle d^2(t)+\varepsilon\rangle}\right)$$

still leads to illegal mass transportation (green boxes in Fig. 3(a)) since the additional masses are not equal

$$\frac{\varepsilon}{\langle s^2(t)+\varepsilon\rangle} \neq \frac{\varepsilon}{\langle d^2(t)+\varepsilon\rangle}.$$

Obviously, the newly created masses in the real and synthetic signals do not match correctly, which could result in an unreasonable optimal transport map and further lead to artifacts in the sensitivity kernel $K_i(\boldsymbol{x})$ (Fig. 3(b)). However, with a simple trick, we can solve the problem of unequal additional masses by modifying the operator as

$$\mathcal{P}_3\big(s(t)\big) = \frac{s^2(t)/\langle s^2(t)\rangle+\varepsilon}{1+t_f\varepsilon}. \tag{3.3}$$

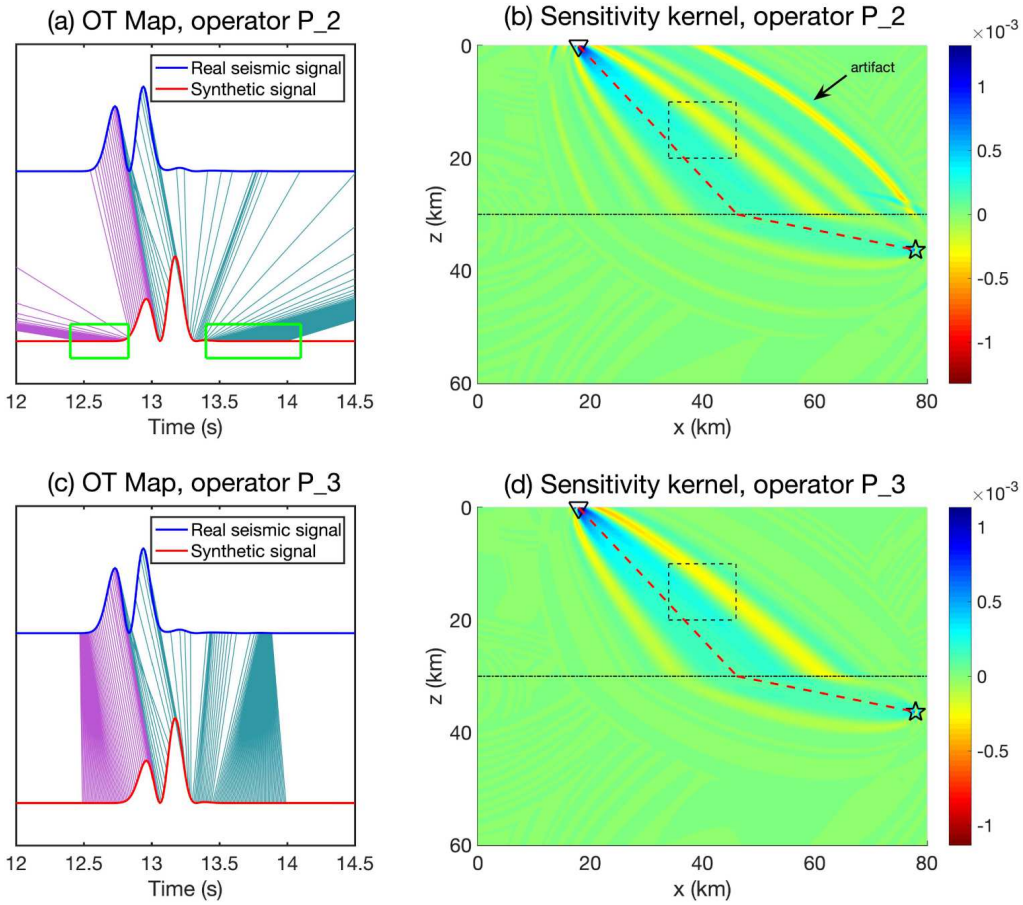

Figure 3: Illustration of the optimal transport maps between the real seismic signal and synthetic signal (a), (c) and the sensitivity kernels (b), (d). The inappropriate operator $\mathcal{P}_2$ in Eq. (3.2) is used for (a) and (b). The newly proposed operator $\mathcal{P}_3$ in Eq. (3.3) is used for (c) and (d). In (a) and (c), the purple and cyan lines represent the mass transportation from the synthetic signal at time t to the real seismic signal at time $T(t)$ according to the optimal transport plan in Eq. (2.19). The different colors correspond to the different time ranges. The illegal mass transportation from the seismic wave to the additional mass caused by the inappropriate operator $\mathcal{P}_2$ is within the green boxes in (a). In (b) and (d), the earthquake and the station receiver are represented by the black star and inverted triangle, respectively. The velocity anomaly is within the black dashed box.

Here, t_f denotes the length of the computational time interval. We can clearly see that regardless of the values of $s(t)$ and $d(t)$, the additional mass is $\varepsilon/(1+t_f\varepsilon)$. As a result, we can avoid all the mentioned troubles. Both the optimal transport map and the sensitivity kernel are satisfactory, as we illustrate in Figs. 3(c) and 3(d).

Remark 3.2. In the squaring scaling, a parameter ε is added to avoid the singularity. It is noted that large ε could destroy the convexity property. On the other hand, numerical singularities occur when ε is too small. In practice, ε is feasible in a relatively large range, e.g., $10^{-4} \sim 10^{-2}$. In the following numerical experiments, we select $\varepsilon = 10^{-3}$.

4. Numerical experiments

In this section, we verify our inversion method based on the quadratic Wasserstein metric with modified squaring scaling by conducting two numerical experiments: (1) the two-layer model [14]; (2) the crustal root model [42]. We use the finite difference method to solve the acoustic wave equation [9, 21, 47]. The perfectly matched layer boundary condition [20] is applied to absorb the outgoing wave. The delta function $\delta(\boldsymbol{x})$ in the source term should be discretized for numerical simulation. Here we choose a simple and effective implementation from [44]

$$\delta_h(x)=\begin{cases}\dfrac{1}{h}\left(1-\dfrac{5}{4}\left|\dfrac{x}{h}\right|^2-\dfrac{35}{12}\left|\dfrac{x}{h}\right|^3+\dfrac{21}{4}\left|\dfrac{x}{h}\right|^4-\dfrac{25}{12}\left|\dfrac{x}{h}\right|^5\right), & |x|\le h,\\ \dfrac{1}{h}\left(-4+\dfrac{75}{4}\left|\dfrac{x}{h}\right|-\dfrac{245}{8}\left|\dfrac{x}{h}\right|^2+\dfrac{545}{24}\left|\dfrac{x}{h}\right|^3-\dfrac{63}{8}\left|\dfrac{x}{h}\right|^4+\dfrac{25}{24}\left|\dfrac{x}{h}\right|^5\right), & h<|x|\le 2h,\\ \dfrac{1}{h}\left(18-\dfrac{153}{4}\left|\dfrac{x}{h}\right|+\dfrac{255}{8}\left|\dfrac{x}{h}\right|^2-\dfrac{313}{24}\left|\dfrac{x}{h}\right|^3+\dfrac{21}{8}\left|\dfrac{x}{h}\right|^4-\dfrac{5}{24}\left|\dfrac{x}{h}\right|^5\right), & 2h<|x|\le 3h,\\ 0, & |x|>3h.\end{cases}$$

Here h is related to the mesh size.

4.1. The two-layer model

The parameter settings can be found in Section 3. We perform the seismic velocity inversion by using the quadratic Wasserstein metric with squaring scaling. The inversion is also performed with the traditional L^2 metric for comparison. To quantitatively compare the results of different methods, we also compute the relative model error

$$\mathrm{RME}=\frac{\displaystyle\int_\Omega |c_k(\boldsymbol{x})-c_T(\boldsymbol{x})|^2\mathrm{d}\boldsymbol{x}}{\displaystyle\int_\Omega |c_0(\boldsymbol{x})-c_T(\boldsymbol{x})|^2\mathrm{d}\boldsymbol{x}},$$

and the relative misfit function

$$\mathrm{RMF}=\frac{\Xi\big(c_k(\boldsymbol{x})\big)}{\Xi\big(c_0(\boldsymbol{x})\big)},$$

where $c_k(\boldsymbol{x})$ indicates the velocity model in the k-th iteration.

In Fig. 4, we present the inversion results of L^2 metric and the quadratic Wasserstein metric. Obviously, the L^2-based inversion could not capture the $+15\%$ high-velocity anomaly (Fig. 4(a)). Although the misfit function decreases in the iteration (Fig. 4(c)), the model error increases (Fig. 4(b)).

In Fig. 4 and Table 1, we also compare the inversion results of the quadratic Wasserstein metric with different operators $\mathcal{P}_2$ and $\mathcal{P}_3$. From the convergent trajectories (Figs. 4(b), 4(c)), we can see the relative model error and the relative misfit function of

Table 1: The two-layer model. Relative model error and relative misfit function of the quadratic Wasserstein with the operators $\mathcal{P}_2$ and $\mathcal{P}_3$ in 20, 40 and 80 iteration steps, respectively.

Iteration Steps	Relative Model Error		Relative Misfit Function	
	W_2 with $\mathcal{P}_2$	W_2 with $\mathcal{P}_3$	W_2 with $\mathcal{P}_2$	W_2 with $\mathcal{P}_3$
20	3.69×10^{-1}	2.15×10^{-1}	4.99×10^{-3}	6.90×10^{-4}
40	2.23×10^{-1}	1.04×10^{-1}	3.61×10^{-4}	9.41×10^{-5}
80	8.35×10^{-2}	3.04×10^{-2}	1.25×10^{-5}	2.75×10^{-6}

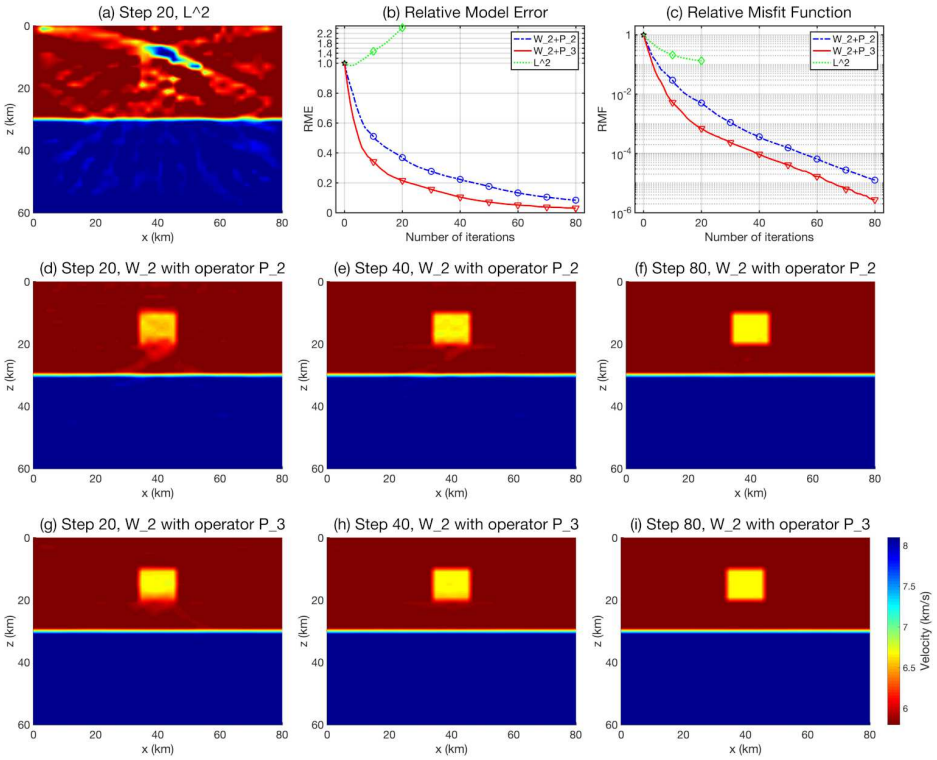

Figure 4: The inversion results of the two-layer model. (a) The result for L^2 metric after 20 steps. (b) The convergent trajectories of the relative model error. (c) The convergent trajectories of the relative misfit function. In (d)-(f) and (g)-(i), we present the results for the quadratic Wasserstein metric with the operators $\mathcal{P}_2$ and $\mathcal{P}_3$, respectively. From left to right, the inversion iteration steps are 20, 40, and 80. All the results are shown in the same color bar.

the operator $\mathcal{P}_3$ are significantly smaller than those of the operator $\mathcal{P}_2$. Quantitatively, we can see from Table 1 that the operator $\mathcal{P}_3$ only needs half of the iteration steps of the operator $\mathcal{P}_2$ to achieve almost the same relative model error and relative misfit function. This significantly reduces the expensive computational cost of the seismic velocity inversion problem. It can be also seen from Figs. 4(d)-4(i) that the velocity inversion results of the operator $\mathcal{P}_3$ are better than those of the operator $\mathcal{P}_2$ at the same iteration steps. The above discussions show that our approach has higher efficiency and generates better inversion results.

4.2. The crustal root model

Let us consider the crustal root model, a kind of subsurface structure usually found along the orogen. This model consists of the two-layered crust divided by the Conrad discontinuity. A dipping and discontinuous Moho interface separates the crust and the mantle. The depiction of these tectonic features helps us better understand the forming of old mountains. In mathematics, we consider this three-layer model in the bounded domain $\Omega = [0, 80\ \mathrm{km}] \times [0, 80\ \mathrm{km}]$. Three layers are divided by the Conrad discontinuity at $20\ \mathrm{km}$ depth and the Moho discontinuity whose location $(x, L(x))$ is formulated with a piecewise function, given by

$$L(x) = \begin{cases} 36 + \dfrac{25}{1600}x^2\ \mathrm{km}, & 0\ \mathrm{km} \le x \le 40\ \mathrm{km}, \\ 36\ \mathrm{km}, & 40\ \mathrm{km} < x \le 80\ \mathrm{km}. \end{cases}$$

The seismic wave speed at each layer refers to the AK135 model [19], generating the real seismic velocity model (Fig. 5(a))

$$c_T(x,z) = \begin{cases} 5.8\ \mathrm{km/s}, & z \le 20\ \mathrm{km}, \\ 6.5\ \mathrm{km/s}, & 20\ \mathrm{km} < z \le L(x), \\ 8.04\ \mathrm{km/s}, & \text{otherwise}. \end{cases}$$

Correspondingly, the initial velocity model (Fig. 5(b)) without crustal root anomaly is

$$c_0(x,z) = \begin{cases} 5.8\ \mathrm{km/s}, & z \le 20\ \mathrm{km}, \\ 6.5\ \mathrm{km/s}, & 20\ \mathrm{km} < z \le 36\ \mathrm{km}, \\ 8.04\ \mathrm{km/s}, & \text{otherwise}. \end{cases}$$

Our goal is to perform the seismic velocity inversion to image this crustal root.

The computational time interval is $[0\ \mathrm{s}, 21\ \mathrm{s}]$. The inversion grid spacing is $2\ \mathrm{km}$ and the number of degrees of freedom amounts to 1600. The space and time steps in the forward simulation are $0.2\ \mathrm{km}$ and $0.01\ \mathrm{s}$, respectively. The dominant frequency of the source time function in (2.7) is $f_0 = 2\ \mathrm{Hz}$. We randomly choose 40 receiver stations deployed on the surface and 80 earthquakes distributed in the study region.

Similar to Section 4.1, we present the inversion results of the L^2 metric and the quadratic Wasserstein metric with the operators $\mathcal{P}_2$ and $\mathcal{P}_3$ in Fig. 6. Obviously, the L^2-based inversion could not capture the crustal root structure as shown in Fig. 6(a). The relative model error and the relative misfit function with respect to different normalization operators are given in Table 2. Correspondingly, the convergent trajectories are output in Figs. 6(b) and 6(c). In Figs. 6(d)-6(i), the inversion results are also presented. According to the above results, we can draw the same conclusions as those in Section 4.1.

Table 2: The crustal root model. Relative model error and relative misfit function of the quadratic Wasserstein with the operators $\mathcal{P}_2$ and $\mathcal{P}_3$ in 40, 80 and 160 iteration steps, respectively.

Iteration Steps	Relative Model Error		Relative Misfit Function	
	W_2 with $\mathcal{P}_2$	W_2 with $\mathcal{P}_3$	W_2 with $\mathcal{P}_2$	W_2 with $\mathcal{P}_3$
40	6.43×10^{-1}	5.59×10^{-1}	5.47×10^{-3}	6.35×10^{-4}
80	5.37×10^{-1}	4.68×10^{-1}	7.83×10^{-4}	1.74×10^{-4}
160	4.32×10^{-1}	3.99×10^{-1}	1.33×10^{-4}	6.11×10^{-5}

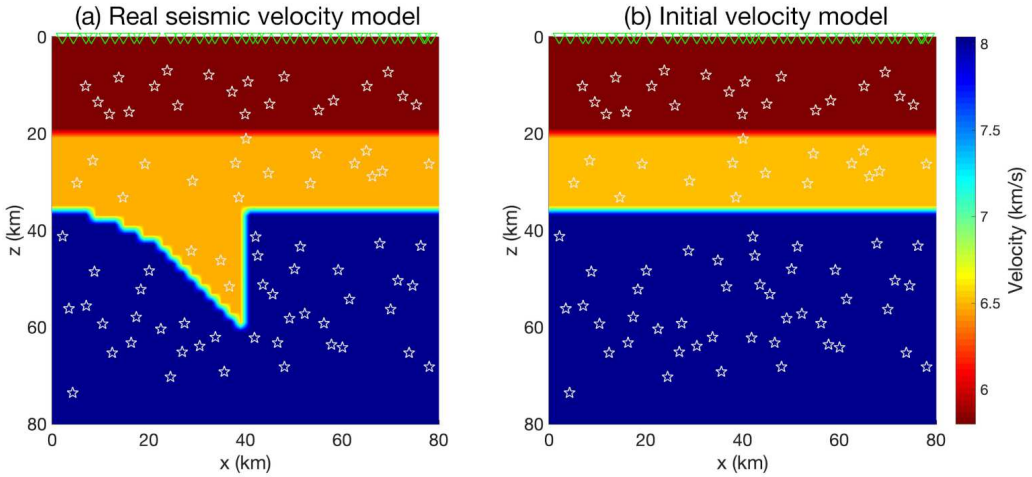

Figure 5: Illustration of the crustal root models. (a) The real seismic velocity model. (b) The initial velocity model. The green inverted triangles and the white stars indicate the receiver stations and the earthquakes, respectively.

5. Conclusion

What we have seen from the above is the solution to the difficult problem of seismic velocity inversion based on the quadratic Wasserstein metric with squaring scaling. The key step is to obtain a reasonable and accurate optimal transport map, which plays a dominant role in constructing the Fréchet derivative. By focusing on the comparison between the same seismic phases and carefully designing the normalization method, we obtain reasonable and accurate optimal transport maps and succeed in inverting for the regional-scale seismic velocity structures using the quadratic Wasserstein metric with modified squaring scaling. The experiment results show that the convergence efficiency and result accuracy are significantly improved. In later works, we would like to combine the above techniques with the double-difference travel-time adjoint tomography [6], which has significant advantages in real seismic data. This may result in a more robust and reliable seismic velocity inversion method. We are currently investigating this interesting topic and hope to report this in an independent publication.

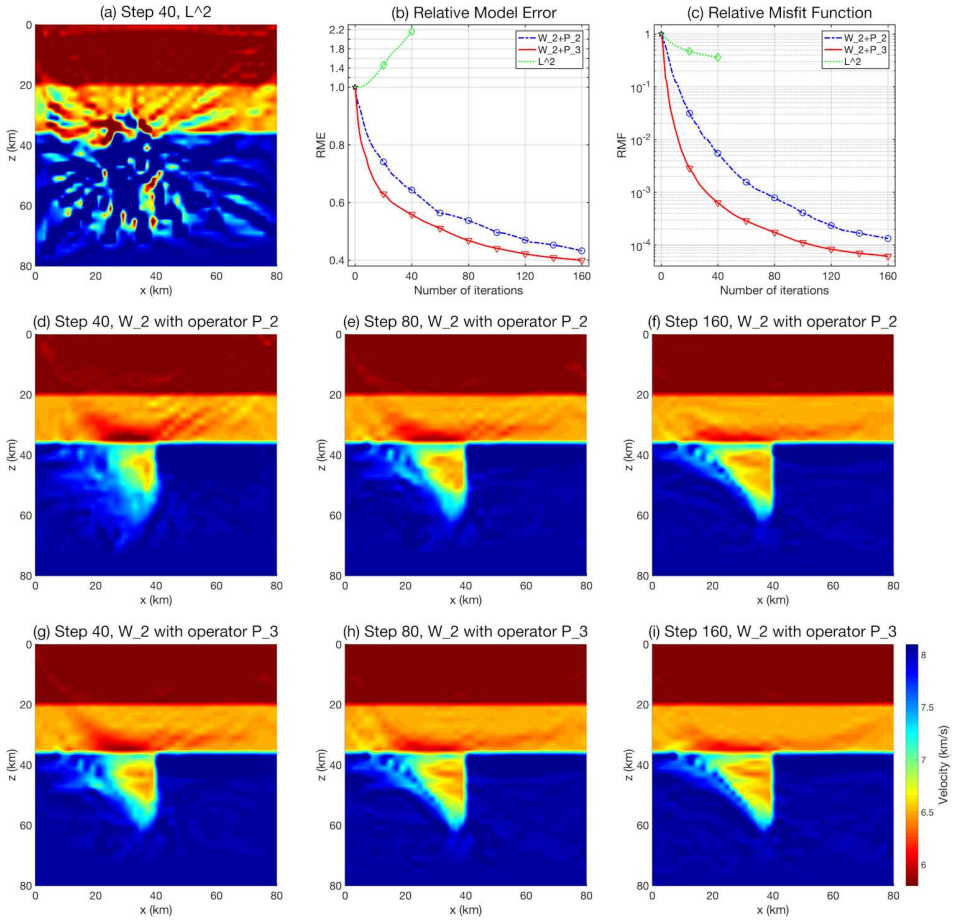

Figure 6: The inversion results of the crustal root model. (a) The result for L^2 metric after 40 steps. (b) The convergent trajectories of the relative model error. (c) The convergent trajectories of the relative misfit function. In (d)-(f) and (g)-(i), we present the results for the quadratic Wasserstein metric with the operators $\mathcal{P}_2$ and $\mathcal{P}_3$, respectively. From left to right, the inversion iteration steps are 40, 80, and 160. All the results are shown in the same color bar.

Acknowledgments

The authors would appreciate the editor and two anonymous reviewers for their constructive suggestions and remarkable comments, which are helpful to improve this article.

This work was supported by the National Natural Science Foundation of China (Grant No. 12271289).

References

[1] K. Aki and P. G. Richards, *Quantitative Seismology: Theory and Methods*. Vol. II, W. H. Freeman and Co., 1980.

[2] E. Bozdağ, J. Trampert, and J. Tromp, *Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements*, Geophys. J. Int. 185(2) (2011), 845–870.
[3] R. Brossier, S. Operto, and J. Virieux, *Velocity model building from seismic reflection data by full-waveform inversion*, Geophys. Prospect. 63(2) (2015), 354–367.
[4] C. Bunks, F. M. Saleck, S. Zaleck, and G. Chavent, *Multiscale seismic waveform inversion*, Geophysics 60(5) (1995), 1457–1473.
[5] B. X. Chen, L. G. Dong, and Y. Z. Liu, *Full waveform inversion method using envelope objective function without low frequency data*, J. Appl. Geophys. 109 (2014), 36–46.
[6] J. Chen, G. X. Chen, H. Wu, J. Y. Yao, and P. Tong, *Adjoint tomography of northeast Japan revealed by common-source double-difference travel-time data*, Seismol. Res. Lett. 93(3) (2022), 1835–1851.
[7] J. Chen, Y. F. Chen, H. Wu, and D. H. Yang, *The quadratic Wasserstein metric for earthquake location*, J. Comput. Phys. 373 (2018), 188–209.
[8] R. Chu, S. Ni, A. Pitarka, and D. V. Helmberger, *Inversion of source parameters for moderate earthquakes using short-period teleseismic P waves*, Pure Appl. Geophys. 171(7) (2014), 1329–1341.
[9] M. A. Dablain, *The application of high-order differencing to the scalar wave equation*, Geophysics 51(1) (1986), 54–66.
[10] B. Engquist and B. D. Froese, *Application of the Wasserstein metric to seismic signals*, Commun. Math. Sci. 12(5) (2014), 979–988.
[11] B. Engquist, B. D. Froese, and Y. N. Yang, *Optimal transport for seismic full waveform inversion*, Commun. Math. Sci. 14(8) (2016), 2309–2330.
[12] B. Engquist and Y. N. Yang, *Seismic imaging and optimal transport*, Commun. Inf. Syst. 19(2) (2019), 95–145.
[13] B. Engquist and Y. N. Yang, *Seismic inversion and the data normalization for optimal transport*, Methods Appl. Anal. 26(2) (2019), 133–148.
[14] B. Engquist and Y. N. Yang, *Optimal transport based seismic inversion: Beyond cycle skipping*, Comm. Pure Appl. Math. 75(10) (2021), 2201–2244.
[15] A. Górszczyk, R. Brossier, and L. Métivier, *Graph-space optimal transport concept for time-domain full-waveform inversion of ocean-bottom seismometer data: Nankai trough velocity structure reconstructed from a 1D model*, J. Geophys. Res. Solid Earth 126(5) (2021), e2020JB021504.
[16] X. Huang, D. Yang, P. Tong, J. Badal, and Q. Liu, *Wave equation-based reflection tomography of the 1992 Landers earthquake area*, Geophys. Res. Lett. 43(5) (2016), 1884–1892.
[17] S.-H. Hung, F. A. Dahlen, and G. Nolet, *Wavefront healing: A banana–doughnut perspective*, Geophys. J. Int. 146(2) (2001), 289–312.
[18] B. L. N. Kennett and E. R. Engdahl, *Traveltimes for global earthquake location and phase identification*, Geophys. J. Int. 105(2) (1991), 429–465.
[19] B. L. N. Kennett, E. R. Engdahl, and R. Buland, *Constraints on seismic velocities in the Earth from traveltimes*, Geophys. J. Int. 122(1) (1995), 108–124.
[20] D. Komatitsch and J. Tromp, *A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation*, Geophys. J. Int. 154(1) (2003), 146–153.
[21] J. S. Li, D. H. Yang, H. Wu, and X. Ma, *A low-dispersive method using the high-order stereo-modelling operator for solving 2-D wave equations*, Geophys. J. Int. 210(3) (2017), 1938–1964.
[22] S. Luo and P. Sava, *A deconvolution-based objective function for wave-equation inversion*,

SEG Technical Program Expanded Abstracts (2011), 2788–2792.
[23] Y. LUO AND G. T. SCHUSTER, *Wave-equation traveltime inversion*, Geophysics 56(5) (1991), 645–653.
[24] R. MADARIAGA, *Seismic source theory*, in: *Treatise on Geophysics*, Elsevier, 2015, 51–71.
[25] A. MAGGI, C. TAPE, M. CHEN, D. CHAO, AND J. TROMP, *An automated time-window selection algorithm for seismic tomography*, Geophys. J. Int. 178(1) (2009), 257–281.
[26] L. MÉTIVIER, A. ALLAIN, R. BROSSIER, Q. MÉRIGOT, E. OUDET, AND J. VIRIEUX, *Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach*, Geophysics 83(5) (2018), R515–R540.
[27] L. MÉTIVIER, R. BROSSIER, Q. MÉRIGOT, E. OUDET, AND J. VIRIEUX, *Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion*, Geophys. J. Int. 205(1) (2016), 345–377.
[28] L. MÉTIVIER, R. BROSSIER, Q. MÉRIGOT, E. OUDET, AND J. VIRIEUX, *An optimal transport approach for seismic tomography: Application to 3D full waveform inversion*, Inverse Problems 32(11) (2016), 115008.
[29] L. MÉTIVIER, R. BROSSIER, J. VIRIEUX, AND S. OPERTO, *Full waveform inversion and the truncated Newton method*, SIAM J. Sci. Comput. 35(2) (2013), B401–B437.
[30] S. M. MOUSAVI, W. L. ELLSWORTH, W. Q. ZHU, L. Y. CHUNG, AND G. C. BEROZA, *Earthquake transformer – an attentive deep-learning model for simultaneous earthquake detection and phase picking*, Nat. Commun. 11(1) (2020), 1–12.
[31] R.-E. PLESSIX, *A review of the adjoint-state method for computing the gradient of a functional with geophysical applications*, Geophys. J. Int. 167(2) (2006), 495–503.
[32] R. G. PRATT, *Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model*, Geophysics 64(3) (1999), 888–901.
[33] L. QIU, J. RAMOS-MARTÍNEZ, A. VALENCIANO, Y. YANG, AND B. ENGQUIST, *Full-waveform inversion with an exponentially encoded optimal-transport norm*, SEG Technical Program Expanded Abstracts (2017), 1286–1290.
[34] N. RAWLINSON, M. SAMBRIDGE, AND J. HAUSER, *Multipathing, reciprocal traveltime fields and raylets*, Geophys. J. Int. 181(2) (2010), 1077–1092.
[35] Z. E. ROSS, M. A. MEIER, AND E. HAUKSSON, *P wave arrival picking and first-motion polarity determination with deep learning*, J. Geophys. Res. Solid Earth 123(6) (2018), 5120–5129.
[36] F. SANTAMBROGIO, *Optimal Transport for Applied Mathematicians*, Birkhäuser, 2015.
[37] C. TAPE, Q. Y. LIU, A. MAGGI, AND J. TROMP, *Seismic tomography of the southern California crust based on spectral-element and adjoint methods*, Geophys. J. Int. 180(1) (2010), 433–462.
[38] J. TROMP, C. TAPE, AND Q. Y. LIU, *Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels*, Geophys. J. Int. 160(1) (2005), 195–216.
[39] C. VILLANI, *Topics in Optimal Transportation*, Graduate Studies in Mathematics, AMS, 2003.
[40] C. VILLANI, *Optimal Transport: Old and New*, Springer, 2008.
[41] J. VIRIEUX AND S. OPERTO, *An overview of full-waveform inversion in exploration geophysics*, Geophysics 74(6) (2009), WCC1–WCC26.
[42] J. WANG, D. H. YANG, H. JING, AND H. WU, *Full waveform inversion based on the ensemble Kalman filter method using uniform sampling without replacement*, Sci. Bull. 64(5) (2019), 321–330.
[43] S. WANG AND H. TKALČIĆ, *Seismic event coda-correlation: Toward global coda-correlation tomography*, J. Geophys. Res. Solid Earth 125 (2020), e2019JB018848.

[44] X. WEN, *High order numerical quadratures to one dimensional delta function integrals*, SIAM J. Sci. Comput. 30(4) (2008), 1825–1846.
[45] S. XU, D. L. WANG, F. CHEN, G. LAMBARÉ, AND Y. ZHANG, *Inversion on reflected seismic wave*, SEG Technical Program Expanded Abstracts (2012), 1–7.
[46] Y. N. YANG AND B. ENGQUIST, *Analysis of optimal transport and related misfit functions in full-waveform inversion*, Geophysics 83(1) (2018), A7–A12.
[47] 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 83(1) (2018), R43–R62.
[48] Ö. YILMAZ, *Engineering Seismology with Applications to Geotechnical Engineering*, Society of Exploration Geophysicists, 2015.
[49] D. T. ZHOU, J. CHEN, H. WU, D. H. YANG, AND L. Y. QIU, *The Wasserstein-Fisher-Rao metric for waveform based earthquake location*, J. Comp. Math. 41(3) (2023), 437–458.
[50] W. ZHOU, R. BROSSIER, S. OPERTO, AND J. VIRIEUX, *Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation*, Geophys. J. Int. 202(3) (2015), 1535–1554.