

# THE WASSERSTEIN-FISHER-RAO METRIC FOR WAVEFORM BASED EARTHQUAKE LOCATION\*

Datong Zhou, Jing Chen, Hao Wu<sup>1)</sup> and Dinghui Yang

Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China

Email: zdt14@mails.tsinghua.edu.cn, jing-che16@mails.tsinghua.edu.cn,

hwu@tsinghua.edu.cn, ydh@mail.tsinghua.edu.cn

Lingyun Qiu

Yau Mathematical Sciences Center, Tsinghua University, Beijing China 100084;

Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China

Email: lyqiu@tsinghua.edu.cn

## Abstract

In this paper, we apply the Wasserstein-Fisher-Rao (WFR) metric from the unbalanced optimal transport theory to the earthquake location problem. Compared with the quadratic Wasserstein ( $W_2$ ) metric from the classical optimal transport theory, the advantage of this method is that it retains the important amplitude information as a new constraint, which avoids the problem of the degeneration of the optimization objective function near the real earthquake hypocenter and origin time. As a result, the deviation of the global minimum of the optimization objective function based on the WFR metric from the true solution can be much smaller than the results based on the  $W_2$  metric when there exists strong data noise. Thus, we develop an accurate earthquake location method under strong data noise. Many numerical experiments verify our conclusions.

*Mathematics subject classification:* 65K10, 86C08, 86A15, 86A22.

*Key words:* The Wasserstein-Fisher-Rao metric, The quadratic Wasserstein metric, Inverse theory, Waveform inversion, Earthquake location.

## 1. Introduction

The optimal transport theory is widely applied in seismology in recent years, leading to more accurate inversion results in the field of geophysical inverse problems, e.g., earthquake location and seismic tomography [11–13, 30, 31, 35, 52, 53]. In these models, appropriate seismic parameters should match the synthetic signals with the observations [36]. From the mathematical point of view, an approximate relationship can be established between seismic parameters  $\mathbf{m}$  and synthetic seismic signals  $\mathbf{d}_{syn}(\mathbf{m})$  by numerically calculating wave equations. Solving this inverse problem requires  $\mathbf{m}$  to minimize the difference between the synthetic  $\mathbf{d}_{syn}(\mathbf{m})$  and the observations  $\mathbf{d}_{obs}$  for a specific metric. The adjoint state method is widely applied to this PDE constrained optimization problem [4–6, 14, 42–44, 50]. The Fréchet gradient of the optimization objective function can be obtained by comparing  $\mathbf{d}_{obs}$  with  $\mathbf{d}_{syn}(\mathbf{m})$ , which is used to update the seismic parameters  $\mathbf{m}$ .

In the past, limited by the computational power, the relationship between parameters  $\mathbf{m}$  and signals  $\mathbf{d}_{syn}(\mathbf{m})$  was established based on the ray theory [17, 46, 51]. Under this high-frequency

\* Received February 11, 2020 / Revised version received May 17, 2021 / Accepted September 26, 2021 /

Published online December 7, 2022 /

<sup>1)</sup> Corresponding author

assumption, some finite frequency phenomena such as wave-front healing and scattering are ignored [42], leading to inaccurate inversion results. With the rapid increase of computational power in recent years, it gradually becomes possible to numerically calculate the wave equation to obtain more accurate synthetic seismic signals  $\mathbf{d}_{syn}(\mathbf{m})$ , which mitigates the bias from the high-frequency assumption and raise the inversion resolution reaching the wavelength scale [32, 36, 42, 44, 54, 56, 57].

However, the traditional  $L_2$  norm based waveform inversion suffers from the cycle-skipping problem [26]. Especially for the earthquake location problem, seismic signals are sensitive to the perturbation of the origin time and the earthquake hypocenter. Thus, under the framework of  $L_2$  norm, this point-to-point comparison between signals might generate numerous local minimums, leading to inaccurate inversion results or excessive iteration steps [4, 12]. Based on the optimal transport theory, the Wasserstein metric provides a new perspective to solve these mentioned problems [4, 11, 12, 14, 30, 31, 37, 45, 52, 53]. By comparing signals globally, the Wasserstein metric defined optimization objective function guarantees better convexity property and mitigates the influence of noise. Thus, reasonably accurate inversion results can be expected when the data is contaminated with high-intensity noise [4]. These accurate inversion results could provide significant guidance for the establishment of the early warning system [38], mineral exploration [9], and the siting of major facilities [3].

The quadratic Wasserstein metric requires mass conservation [45]. Thus, the normalization process is mandatory for seismic signals, which becomes an essential limitation. For the earthquake location problem, the amplitude of seismograms provides necessary constraints to the origin time and the distance from the hypocenter to the receiver. Simply normalizing the signals would lead to a nearly flat optimization objective function along a certain direction due to the trade-off between the origin time and the distance. Thus, the minimum point of the optimization objective function may deviate a lot even under the small magnitude of data noise, which leads to low accurate location results, see Examples 3.1 for illustration.

The Wasserstein-Fish-Rao (WFR) metric is a newly developed optimal transport metric and has attracted much attention [7, 8]. This metric is an interpolation between the quadratic Wasserstein metric and the Fisher-Rao metric. From the fluid dynamics point of view [1], this new metric introduces a source term in the continuity equation, allowing the direct comparison between two signals with different total integrals. Thus this metric is also called unbalance optimal transport metric [16, 18, 25, 34], which has been successfully applied in various fields [15, 48, 55]. Benefited from the above features, the normalization of the seismic signals is no longer required. Therefore, the important amplitude information is retained based on the WFR metric, improving the local convexity and avoiding the degeneracy of the optimization objective function near the global minimum point.

**Remark 1.1.** The Kantorovich-Rubinstein (KR) norm [30, 31] does not require the signals to have the same integral. However, the convexity of the optimization objective functions defined with the KR norm may not be guaranteed for the earthquake location problems. For more details, we refer to see [4], especially Figs. 2-3 for illustration.

**Remark 1.2.** In [4], we clearly see the superiority of the quadratic Wasserstein metric to the  $L_2$  metric. So in this paper, we prefer to focus on the comparison between the new WFR metric and the quadratic Wasserstein metric.

In this paper, we introduce the WFR metric to the earthquake location problem. It is a significant extension of optimal transport theory in the application to the geophysical inverse

problem. Compared to the classical Wasserstein metric, the normalization process is no longer compulsory, indicating the amplitude of seismic signals could provide significant constraints to the inversion. Figs. 3.3-3.6 illustrate better local convexity and non-degeneracy for the WFR metric, which means less sensitivity to strong data noise. Thus, we can expect more accurate results for the WFR metric based inversion method.

This paper is organized as follows. In Section 2, the earthquake location problem coupled with the adjoint state method is reviewed. We then introduce the WFR metric and apply it to the earthquake location problem in Section 3. In Section 4, the numerical experiments are provided to demonstrate the effectiveness and efficiency of the new method. Finally, we make some conclusive remarks in Section 5.

## 2. The Earthquake Location Problem

We now review the earthquake location problem and the adjoint state method. From the mathematical point of view, the waveform based earthquake location problem can be formulated as an optimization model with PDE constraints [4, 5, 50]

$$(\boldsymbol{\xi}_T, \tau_T) = \underset{\boldsymbol{\xi}, \tau}{\operatorname{argmin}} \sum_r \chi_r(\boldsymbol{\xi}, \tau). \quad (2.1)$$

Here  $(\boldsymbol{\xi}_T, \tau_T)$  are the true earthquake location and origin time,  $r$  denotes the index of receivers, and the corresponding misfit function  $\chi_r(\boldsymbol{\xi}, \tau)$  with respect to the earthquake location  $\boldsymbol{\xi}$  and origin time  $\tau$  is defined as

$$\chi_r(\boldsymbol{\xi}, \tau) = \mathcal{D}(d_r(t), s(\boldsymbol{\eta}_r, t)) \triangleq \mathcal{D}_r(s). \quad (2.2)$$

In the above equation,  $d_r(t)$  and  $s(\mathbf{x}, t)$  are the observed and synthetic earthquake signals, respectively.  $\mathcal{D}$  measures the distance between them, which will be specified later. For simplification, we will write the Fréchet derivative of  $\mathcal{D}_r(s)$  to  $s(t)$  as  $\nabla_s \mathcal{D}(t)$ . The two wavefields,

$$d_r(t) = u(\boldsymbol{\eta}_r, t; \boldsymbol{\xi}_T, \tau_T), \quad s(\mathbf{x}, t) = u(\mathbf{x}, t; \boldsymbol{\xi}, \tau), \quad (2.3)$$

can be regarded as the solutions to the following scalar acoustic wave equation with initial-boundary conditions

$$\frac{\partial^2 u(\mathbf{x}, t; \boldsymbol{\xi}, \tau)}{\partial t^2} = \nabla \cdot (c^2(\mathbf{x}) \nabla u(\mathbf{x}, t; \boldsymbol{\xi}, \tau)) + R(t - \tau) \delta(\mathbf{x} - \boldsymbol{\xi}), \quad \mathbf{x}, \boldsymbol{\xi} \in \Omega, \quad (2.4)$$

$$u(\mathbf{x}, 0; \boldsymbol{\xi}, \tau) = \partial_t u(\mathbf{x}, 0; \boldsymbol{\xi}, \tau) = 0, \quad \mathbf{x} \in \Omega, \quad (2.5)$$

$$\mathbf{n} \cdot (c^2(\mathbf{x}) \nabla u(\mathbf{x}, t; \boldsymbol{\xi}, \tau)) = 0, \quad \mathbf{x} \in \partial\Omega. \quad (2.6)$$

Here  $c(\mathbf{x})$  is the acoustic wave speed, and  $\boldsymbol{\eta}_r$  denotes the location of the  $r$ -th receiver. The point source  $\delta(\mathbf{x} - \boldsymbol{\xi})$  is used to model the seismic rupture inside the simulation domain  $\Omega \subset \mathbb{R}^d$ . This assumption is reasonable since the scale of seismic rupture is much smaller than that of the seismic wave [27]. The seismic source is simplified modeled in the form of Ricker wavelet

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

in which  $f_0$  is the dominant frequency and  $A$  is the indication parameter of amplitude. On the boundary  $\partial\Omega$ , the reflection condition (2.6) is used for simplification and  $\mathbf{n}$  is the outward unit normal vector to the domain  $\Omega$ . We can also easily consider the other boundary conditions, e.g., the perfectly matched layer absorbing boundary condition [19].

## 2.1. The adjoint method

To solve this optimization problem (2.1), it is necessary to obtain the gradient of the misfit function  $\chi_r(\xi, \tau)$  in (2.2) with respect to the earthquake hypocenter  $\xi$  and the origin time  $\tau$ .

For the perturbation of the misfit  $\delta\chi(\xi, \tau)$  resulted from the perturbation of earthquake hypocenter  $\delta\xi$  and the origin time  $\delta\tau$ , we assume it has the formulation of

$$\begin{aligned}\delta\chi_r &= \mathcal{D}(d_r(t), s(\eta_r, t) + \delta s(\eta_r, t)) - \mathcal{D}(d_r(t), s(\eta_r, t)) = \mathcal{D}_r(s + \delta s) - \mathcal{D}_r(s) \\ &\approx \langle \nabla_s \mathcal{D}_r(t), \delta s(\eta_r, t) \rangle = \int_0^T \nabla_s \mathcal{D}_r(t) \delta s(\eta_r, t) dt.\end{aligned}\quad (2.8)$$

Next, following the similar manner in Section 3.1 of [4], the relation between  $\delta\chi_r$  with  $\delta\xi$  and  $\delta\tau$  can be obtained as

$$\delta\chi_r = K_r^\xi \cdot \delta\xi + K_r^\tau \delta\tau, \quad (2.9)$$

$$K_r^\xi = \int_0^T R(t - \tau) \nabla w_r(\xi, t) dt, \quad (2.10)$$

$$K_r^\tau = - \int_0^T R'(t - \tau) w_r(\xi, t) dt, \quad (2.11)$$

where the adjoint wavefield  $w_r(\xi, t)$  satisfies

$$\frac{\partial^2 w_r(\mathbf{x}, t)}{\partial t^2} = \nabla \cdot (c^2(\mathbf{x}) \nabla w_r(\mathbf{x}, t)) + \nabla_s \mathcal{D}_r(t) \delta(\mathbf{x} - \eta_r), \quad \mathbf{x} \in \Omega, \quad (2.12)$$

$$w_r(\mathbf{x}, T) = \partial_t w_r(\mathbf{x}, T) = 0, \quad \mathbf{x} \in \Omega, \quad (2.13)$$

$$\mathbf{n} \cdot (c^2(\mathbf{x}) \nabla w_r(\mathbf{x}, t)) = 0, \quad \mathbf{x} \in \partial\Omega. \quad (2.14)$$

## 3. The Wasserstein-Fisher-Rao Metric

Consider two non-negative distributions  $\mu$  and  $\nu$  belong to the Borel measures denoted by  $\mathcal{M}(\Omega)$  on a compact convex spatial domain  $\Omega \subset \mathbb{R}^n$  (e.g., Section 1.5 in [7]). The WFR metric with the interpolating parameter  $0 < \gamma < \infty$  is defined by solving a minimizing problem [7]:

$$\text{WFR}_\gamma(\mu, \nu) = \left( \inf_{\rho, v, \alpha} \int_0^1 \int_\Omega \left( \frac{1}{2} |v(t, \mathbf{x})|^2 + \frac{\gamma^2}{2} \alpha(t, \mathbf{x})^2 \right) \rho(t, \mathbf{x}) d\mathbf{x} dt \right)^{\frac{1}{2}}, \quad (3.1)$$

under the constraint that the triplet  $(\rho, v, \alpha)$  satisfies the following continuity equation:

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho v) = \rho \alpha, \\ \rho(0, \cdot) = \mu, \quad \rho(1, \cdot) = \nu. \end{cases} \quad (3.2)$$

Here  $\rho$  is arbitrary time-dependent density,  $v$  is an arbitrary velocity field that stands for the movement of mass, and  $\alpha$  is an arbitrary scalar field associated with the creation and destruction of mass. This kind of metric was defined and studied simultaneously and independently in [7, 16, 18, 24], with quite different approaches. In this paper, we mainly follow the approach of [7].

For  $|\mu| = |\nu|$  and  $\gamma \rightarrow \infty$ , the source term  $\rho\alpha$  in the continuity equation is depleted (i.e.  $\alpha \equiv 0$ ) and the metric degenerates to the dynamical formulation of Wasserstein metric of Benamou and Brenier in [1]. Developing the numerical method of the classical optimal

transport problems is still a very active topic recently, and several numerical methods are suggested [1,2,12,21,23,28,29,33,47]. Some of these methods can be extended to the generalized problems introduced in (3.1)-(3.2).

Like the classical Wasserstein metric, the minimizing problem of the WFR metric has a static primal form (e.g., [7,8])

$$\text{WFR}_\gamma^2(\mu, \nu) = \inf_{\lambda(\mathbf{x}, \mathbf{y})} \left\{ \int c(\mathbf{x}, \mathbf{y}) d\lambda(\mathbf{x}, \mathbf{y}) + KL(\lambda_1 | \mu) + KL(\lambda_2 | \nu) : \right. \\ \left. \int_\Omega \lambda(\mathbf{x}, \mathbf{y}) d\mathbf{y} = \lambda_1(\mathbf{x}), \int_\Omega \lambda(\mathbf{x}, \mathbf{y}) d\mathbf{x} = \lambda_2(\mathbf{y}) \right\}, \quad (3.3)$$

where

$$c(\mathbf{x}, \mathbf{y}) = -\log \left( \cos_+^2 \left( \frac{\|\mathbf{x}_i - \mathbf{y}_j\|}{2\gamma} \right) \right), \quad \cos_+(z) = \begin{cases} \cos(z) & \text{for } z < \frac{\pi}{2}, \\ 0, & \text{for } z \geq \frac{\pi}{2}, \end{cases} \quad (3.4)$$

$$KL(\rho | \mu) = \int \varphi_{\text{KL}} \left( \frac{d\rho}{d\mu} \right) d\mu, \quad \varphi_{\text{KL}}(s) = \begin{cases} s \log s - s + 1, & \text{for } s > 0, \\ 1, & \text{for } s = 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

In order to numerically calculate this metric, its dual form can be written as:

$$\text{WFR}_\gamma^2(\mu, \nu) = 2\gamma^2 \sup_{\phi, \psi} \left\{ \int (1 - e^{-\phi}) d\mu + \int (1 - e^{-\psi}) d\nu : \phi(\mathbf{x}) + \psi(\mathbf{y}) \leq c(\mathbf{x}, \mathbf{y}) \right\}. \quad (3.5)$$

In the numerical realization, these two distributions are assumed to be atomic distributions,

$$\mu = \sum_{i=1}^N \mu_i \delta_{\mathbf{x}_i}, \quad \nu = \sum_{j=1}^M \nu_j \delta_{\mathbf{y}_j}, \quad \mu_i \geq 0, \quad \nu_j \geq 0. \quad (3.6)$$

In this discrete setting, (3.5) can be reformulated as

$$\text{WFR}_\gamma^2 \left( \sum_{i=1}^N \mu_i \delta_{\mathbf{x}_i}, \sum_{j=1}^M \nu_j \delta_{\mathbf{y}_j} \right) \\ = 2\gamma^2 \sup_{\phi_i, \psi_j} \left\{ \sum_i (1 - e^{-\phi_i}) \mu_i + \sum_j (1 - e^{-\psi_j}) \nu_j : \phi_i + \psi_j \leq c(\mathbf{x}_i, \mathbf{y}_j) \right\}. \quad (3.7)$$

To solve this constrained optimization problem directly is difficult in numerical realization. Instead, an entropy regularization is introduced to obtain a free-constrained optimization problem:

$$2\gamma^2 \sup_{\phi_i, \psi_j} \left\{ \sum_i (1 - e^{-\phi_i}) \mu_i + \sum_j (1 - e^{-\psi_j}) \nu_j + \epsilon \sum_{ij} \left( 1 - e^{\frac{\phi_i + \psi_j}{\epsilon}} \right) e^{-\frac{c_{ij}}{\epsilon}} \right\}, \quad (3.8)$$

with  $c_{ij} = c(\mathbf{x}_i, \mathbf{y}_j)$ . This problem is strictly convex by introducing the regularization term. Moreover, the solution of the regularized problem (3.8) is unique and should converge to the original problem (3.7) as  $\epsilon \rightarrow 0$  [8].

The regularized problem (3.8) can be easily solved by the iterative proportional fitting procedure [2, 8, 37]. The global maximum of this problem can be obtained by alternatively maximizing the optimization objective function with respect to  $\phi_i$  and  $\psi_j$ , i.e.

$$\begin{cases} \phi^{(\ell+1)} = \arg \max_{\phi} \sum_i \left( (1 - e^{-\phi_i}) \mu_i - \epsilon e^{\phi_i/\epsilon} \sum_j e^{(\psi_j^{(\ell)} - c_{ij})/\epsilon} \right), \\ \psi^{(\ell+1)} = \arg \max_{\psi} \sum_j \left( (1 - e^{-\psi_j}) \nu_j - \epsilon e^{\psi_j/\epsilon} \sum_i e^{(\phi_i^{(\ell+1)} - c_{ij})/\epsilon} \right). \end{cases} \quad (3.9)$$

These maximization steps can be done by pointwise computation:

$$\begin{cases} \tilde{\phi}_i^{(\ell+1)} = \left( \mu_i / \left( \sum_j e^{-c_{ij}/\epsilon} \tilde{\psi}_j^{(\ell)} \right) \right)^{1/(1+\epsilon)}, \\ \tilde{\psi}_j^{(\ell+1)} = \left( \nu_j / \left( \sum_i e^{-c_{ij}/\epsilon} \tilde{\phi}_i^{(\ell+1)} \right) \right)^{1/(1+\epsilon)}, \end{cases}$$

in which

$$\tilde{\phi}_i^{(\ell)} = e^{\phi_i^{(\ell)}/\epsilon}, \quad \tilde{\psi}_j^{(\ell)} = e^{\psi_j^{(\ell)}/\epsilon}.$$

This is how the WFR metric (3.7) is solved numerically in practice [8]. In [39], there is a detailed discussion on the techniques arising in the implementation of the algorithm.

**Remark 3.1.** The parameter  $\gamma$  controls the interpolation between two terms of the minimization problem (3.1). As suggested in [7], the larger  $\gamma$  is, the lower amount of mass will be created or removed, and the WFR metric's behavior will be more like the classical Wasserstein metric. On the other hand, as  $\gamma$  vanishes, the behavior of the WFR metric will approach to the Fisher-Rao metric. Detailed effects of  $\gamma$  on the WFR metric are discussed via several numerical experiments in Example 3.1.

**Remark 3.2.** The selection of the parameter  $\epsilon$  appearing in (3.8) is still an important and open problem (e.g., the end of Section 4.4.2 in [8]). As we know, the smaller  $\epsilon$  implies more accurate calculation of the WFR metric but leads to more iterative steps. Several heuristic selection criteria are proposed to accelerate the convergence of Sinkhorn iteration, e.g., the first paragraph of Section 4.4.2 in [8], the second to last paragraph of Section 3 in [20], and the last paragraph of Section 4 in [40]. In this paper, we follow their ideas to start the Sinkhorn iteration from  $\epsilon = 1$  and gradually decrease it through the iteration. Detailed implementation and analysis are discussed in Section 4.4.2 of [8].

### 3.1. Application to the Earthquake Location problems

We now turn to the application of the WFR metric to the earthquake location problem. This can be simply done by defining the distance  $\mathcal{D}$  in (2.2) using the WFR metric (3.1)-(3.2), and the optimization objective function is formulated as

$$\Theta(\xi, \tau) = \sum_r \chi_r(\xi, \tau) = \sum_r \mathcal{D}(d_r(t), s(\eta_r, t)) = \sum_r \text{WFR}_{\gamma}^2(d_r(t)^2, s(\eta_r, t)^2). \quad (3.10)$$

In the above equation, we square the seismic data to ensure the non-negativeness. Equations (2.1)–(2.7), (3.1)–(3.2) and (3.10) provide the mathematical model to the waveform based earthquake location problems with the WFR metric. To obtain the gradient of the misfit

function  $\chi_r(\xi, \tau)$  with respect to the earthquake hypocenter  $\xi$  and the origin time  $\tau$ , we firstly define

$$F(\phi, \psi, f, g) = 2\gamma^2 \left( \int_0^T \left(1 - e^{-\phi(t)}\right) f^2(t) dt + \int_0^T \left(1 - e^{-\psi(t)}\right) g^2(t) dt \right),$$

$$\Pi = \{(\phi, \psi) : \phi(t_1) + \psi(t_2) \leq c(t_1, t_2), \quad \forall t_1, t_2 \in [0, T]\},$$

according to (3.5) and (3.10). Here we have

$$\text{WFR}_\gamma^2(f(t)^2, g(t)^2) = \sup_{(\phi, \psi) \in \Pi} F(\phi, \psi, f, g) = F(\phi^*, \psi^*, f, g),$$

where  $\phi^*(t)$  and  $\psi^*(t)$  are specific functions that ensure the functional  $F(\phi, \psi, f, g)$  to reach the supreme. Through a simple discussion, we can get

$$\nabla_g \text{WFR}_\gamma^2(f^2, g^2) = \partial_g F(\phi^*, \psi^*, f, g),$$

thus the perturbation of the WFR metric with respect to  $\delta g$  can be written as

$$\begin{aligned} & \text{WFR}_\gamma^2(f^2, (g + \delta g)^2) - \text{WFR}_\gamma^2(f^2, g^2) \\ &= \int_0^T 4\gamma^2 \left(1 - e^{-\psi^*(t)}\right) g(t) \delta g(t) dt + O(\delta g(t)^2). \end{aligned} \quad (3.11)$$

Ignoring the second-order term in (3.11), this formulation exactly satisfies the assumption of (2.8). Thus, the gradient can be obtained from (2.9)–(2.11) with  $\nabla_s \mathcal{D}_r(t)$  in (2.12) specified as

$$\nabla_s \mathcal{D}_r(t) = 4\gamma^2 (1 - e^{-\psi^*(t)}) s(\eta_r, t) \quad (3.12)$$

### 3.2. Advantages of the WFR metric based optimization objective function

As we have already discussed, the main difference between the WFR metric and the classical quadratic Wasserstein metric [4] is that the WFR metric does not need to normalize the seismic signal, thus retaining the important amplitude information. This amplitude information provides necessary constraints to the earthquake location. In the following example, we are going to illustrate a situation that the optimization objective functions induced by the quadratic Wasserstein metric are degenerate near the minimum. As a consequence, even small noise on the signal would make the global minimum of optimization objective function far from the earthquake hypocenter. By introducing the source term, the WFR metric allows the local, unbalanced comparison between signals. This relieves the degeneracy of the optimization objective function, making the global minimum closed to the actual earthquake hypocenter.

We are aware that the global convexity of optimization objective function induced by the WFR metric is not as good as the quadratic Wasserstein metric since the WFR metric introduces the local Fish-Rao metric. Nevertheless, in the neighborhood of the minimum, good convexity property is still conserved for the WFR metric. Its optimization objective function is not as degenerate as that induced by the classical quadratic Wasserstein metric. As a consequence, less deviation of global minimum under larger magnitude of data noise can still be expected by using the WFR metric based optimization objective function. Now, we perform a numerical example to illustrate this perspective.

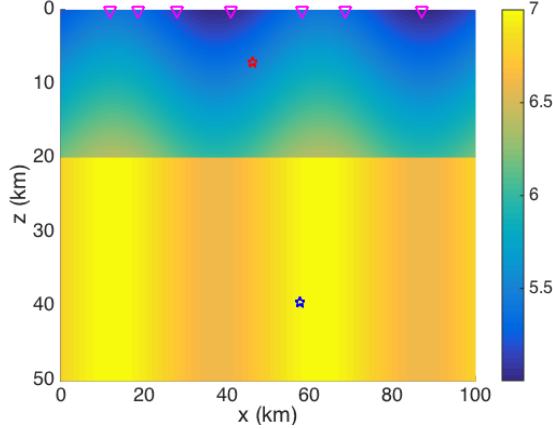


Fig. 3.1. Illustration of two-layer model in Example 3.1. The inverted magenta triangles indicate the receivers. The red and blue pentagrams indicate the earthquake hypocenter above and below the Conrad discontinuity, respectively.

Table 3.1: The two-layer model in Example 3.1: the horizontal positions of receivers, with unit ‘km’.

$r$	1	2	3	4	5	6	7
$x_r$	11.74	18.59	27.88	41.11	58.23	68.50	87.01

**Example 3.1.** This is a 2-D two-layer model in the bounded region  $\Omega = [0, 100 \text{ km}] \times [0, 50 \text{ km}]$  with a mesh size of  $\Delta x = \Delta z = 0.2 \text{ km}$  and time step of  $\Delta t = 0.01 \text{ s}$ . The Conrad discontinuity is located at a depth of 20 km from the Earth’s surface, and the wave speed  $c(x, z)$  is

$$c(x, z) = \begin{cases} 5.2 + 0.05z + 0.2 \sin(\pi x/25), & z \leq 20 \text{ km}, \\ 6.8 + 0.2 \sin(\pi x/25), & z > 20 \text{ km}. \end{cases}$$

The unit is ‘km/s’. We randomly set up 7 receivers on the Earth’s surface ( $z = 0 \text{ km}$ ). Their horizontal positions are listed in Table 3.1, see also Fig. 3.1 for the illustration.

Here we consider two different earthquake hypocenters, one is above the Conrad discontinuity

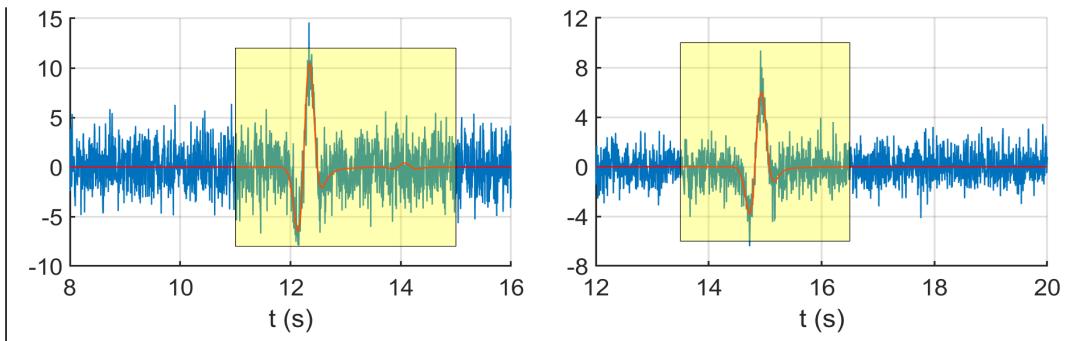


Fig. 3.2. Illustration of signal with noise in the two-layer model. The signal with noise  $d_r(t)$  (blue line) and the noise-free signal  $u(\boldsymbol{\eta}_r, t; \boldsymbol{\xi}_T, \tau_T)$  (red line) for receiver  $r = 1$ . The horizontal axis is the time  $t$ . Yellow intervals mean the time window for the quadratic Wasserstein metric, which is not necessary for the WFR metric. Left: parameters (i) earthquake hypocenter above the Conrad discontinuity; Right: parameters (ii) earthquake hypocenter below the Conrad discontinuity.

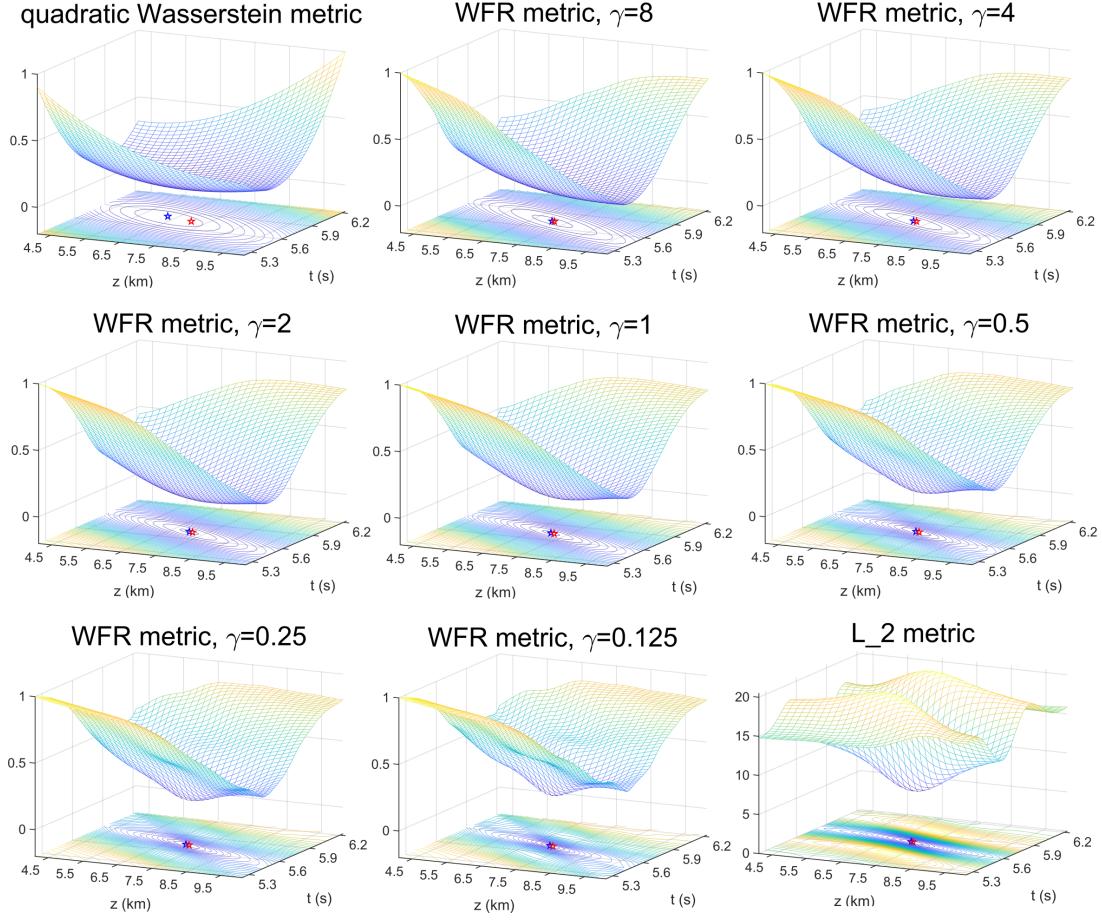


Fig. 3.3. The optimization objective function of the two-layer model with different metrics in Example 3.1 for cross-section at  $x = 46.23$  km of case (i). The blue pentagram denotes the minimum point of the optimization objective function, while the red pentagram denotes the real earthquake hypocenter.

nuity and the other is below the Conrad discontinuity:

$$(i) \quad \xi_T^a = (46.23 \text{ km}, 7.12 \text{ km}), \quad \tau_T^a = 5.73 \text{ s}, \\ (ii) \quad \xi_T^b = (57.60 \text{ km}, 39.36 \text{ km}), \quad \tau_T^b = 5.18 \text{ s}.$$

The dominant frequency of the earthquake is  $f_0 = 2$  Hz, and the amplitude indication parameter is  $A = 15000$ . The real earthquake signals are contaminated with noise

$$d_r(t) = u(\boldsymbol{\eta}_r, t; \boldsymbol{\xi}_T, \tau_T) + N_r(t). \quad (3.13)$$

Here  $N_r(t)$  is subject to the normal distribution with mean  $\mu = 0$  and the standard deviation

$$\sigma = R \times \max_t |u(\boldsymbol{\eta}_r, t; \boldsymbol{\xi}_T, \tau_T)|, \quad (3.14)$$

in which the ratio  $R = 20\%$ . These signals are illustrated in Fig. 3.2. For the quadratic Wasserstein metric and the  $L_2$  metric, a time window (yellow interval) that contains the main part of  $u(\boldsymbol{\eta}_r, t; \boldsymbol{\xi}_T, \tau_T)$  is determined to reduce the impact of noise. In contrast, there is no

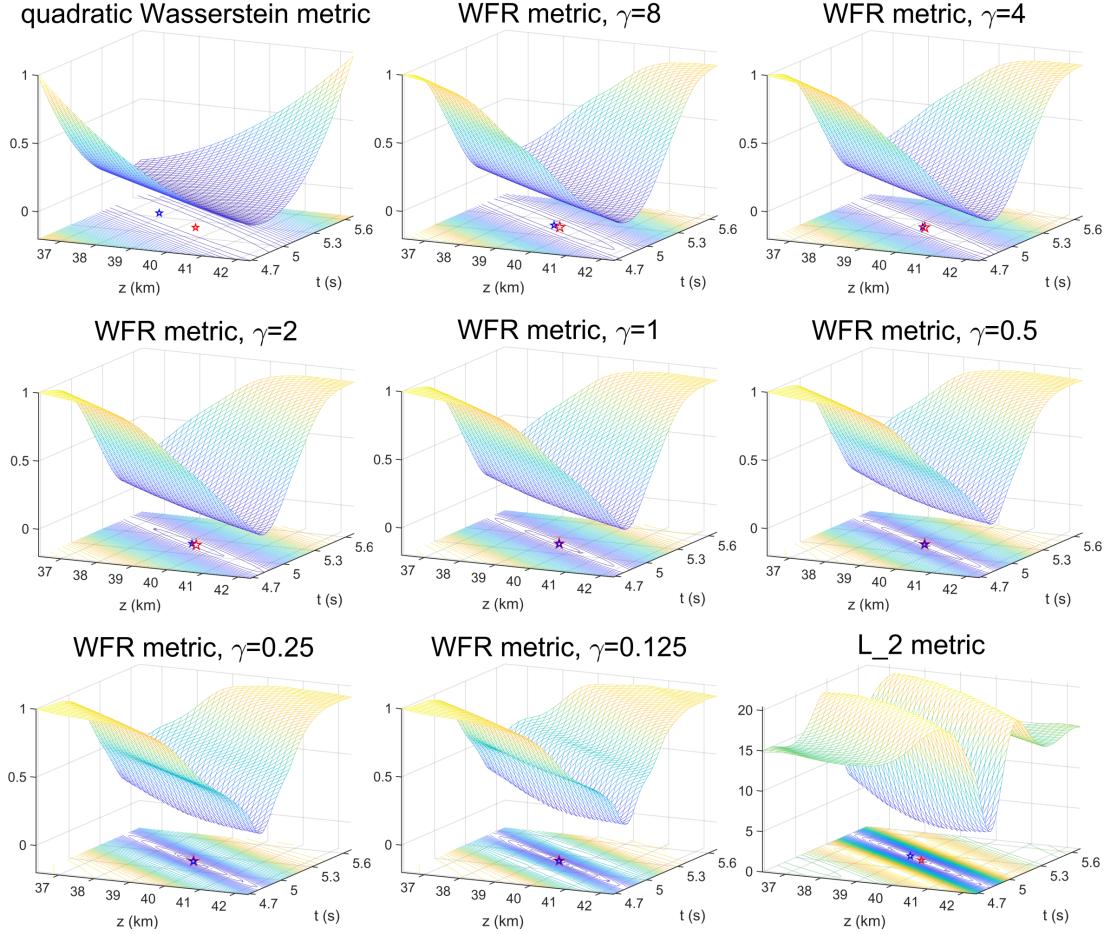


Fig. 3.4. The optimization objective function of the two-layer model with different metrics in Example 3.1 for cross-section at  $x = 57.60$  km of case (ii). The blue pentagram denotes the minimum point of the optimization objective function, while the red pentagram denotes the real earthquake hypocenter.

need for the time window under the WFR metric since the impact of noise is less significant. It implies better robustness for the WFR metric when dealing with noised seismic signals.

We output the cross-section of the optimization objective function defined by the quadratic Wasserstein ( $W_2$ ) metric, WFR metrics with different  $\gamma$ , and the  $L_2$  metric in Figs. 3.3 and 3.4. It is observed that the optimization objective function based on the  $L_2$  metric suffers from the well-known cycle skipping problem, leading to poor convexity and making it more difficult to converge to the real earthquake hypocenter. Affected by the noise, there occur deviations of the minimum point of the optimization objective functions. Table 3.2 presents the offsets between the minimum point and the real earthquake hypocenter for different metrics. This comparison means much higher accuracy for the earthquake location based on the WFR metric than the quadratic Wasserstein metric, even for a wide range of  $\gamma$ .

Next, we test the effects of noise using 100 experiments with white noise. We fix the noise ratio  $R$  at 20% and choose different random seeds. For each experiment, we directly calculate all possible values of the optimization objective function through a brute force algorithm and arrive at the optimal point. In each scenario displayed in Figs. 3.5 and 3.6, all 100 optimal solution

Table 3.2: The two-layer model in Example 3.1: the offset between the minimum point and the real earthquake hypocenter for the quadratic Wasserstein ( $W_2$ ) metric, WFR metrics with different  $\gamma$ , and the  $L_2$  metric, with unit ‘km’. The bold and underlined numbers are the best and second best results, respectively.

	(i) above the Conrad	(ii) below the Conrad
$W_2$ metric	0.8208	1.5898
WFR metric, $\gamma = 8$	<u>0.0699</u>	0.2318
$\gamma = 4$	0.0945	0.0728
$\gamma = 2$	0.1031	0.1685
$\gamma = 1$	0.1025	0.0159
$\gamma = 0.5$	0.1040	0.0015
$\gamma = 0.25$	0.1052	<b>0.0001</b>
$\gamma = 0.125$	0.0995	<u>0.0008</u>
$L_2$ metric	<b>0.0629</b>	0.4747

Table 3.3: The two-layer model in Example 3.1: the average (AVE) and maximal (MAX) distance between the minimum point and the real earthquake hypocenter for 100 tests, with unit ‘km’. We compare the results of the quadratic Wasserstein ( $W_2$ ) metric, the WFR metric, and the  $L_2$  metric. The bold and underlined numbers are the best and second best results, respectively.

	(i) above the Conrad		(ii) below the Conrad	
	AVE	MAX	AVE	MAX
$W_2$ metric	0.5211	2.1596	0.8553	2.5962
WFR metric, $\gamma = 8$	0.0792	<u>0.2837</u>	0.2434	0.8137
$\gamma = 4$	0.0594	0.2626	0.1579	0.6078
$\gamma = 2$	0.0465	0.1842	0.1181	0.3354
$\gamma = 1$	0.0390	0.1244	0.0943	0.3050
$\gamma = 0.5$	<b>0.0359</b>	<u>0.1173</u>	0.0943	<b>0.2709</b>
$\gamma = 0.25$	0.0383	0.1283	<b>0.0686</b>	0.3108
$\gamma = 0.125$	<u>0.0369</u>	<b>0.1155</b>	<u>0.0715</u>	<u>0.2770</u>
$L_2$ metric	0.0454	0.1360	0.3785	1.1713

points are presented with the background showing the contour of the noise-free optimization objective function. The primary purpose of these experiments is to study the location bias in the statistical sense. The averaged and maximum distances between the optimal solution point and the real earthquake hypocenter are displayed in Table 3.3. These statistical results indicate the deviations of minimum points of WFR metrics with different  $\gamma$  are similar and much better than the quadratic Wasserstein metric. Though the offset for the  $L_2$  metric is slightly better than the  $W_2$  metric and even comparable with the WFR metric for shallow earthquakes above Conrad. However, the global minimum for  $L_2$  metric deviates much more than the WFR metric when the earthquake occurs deeply. These experiment results verify our previous perspective. Higher accuracy for the earthquake location can be achieved based on the WFR metric. In addition, the location accuracy is less sensitive to the selection of  $\gamma$  in a certain interval.

The above example illustrates that the WFR metric can be used to obtain more accurate earthquake location results than the quadratic Wasserstein metric under the same magnitude of noise. Compared to the  $L_2$  metric, both higher location accuracy and better convexity of the optimization objective function can be guaranteed by the WFR metric. Thus, we believe

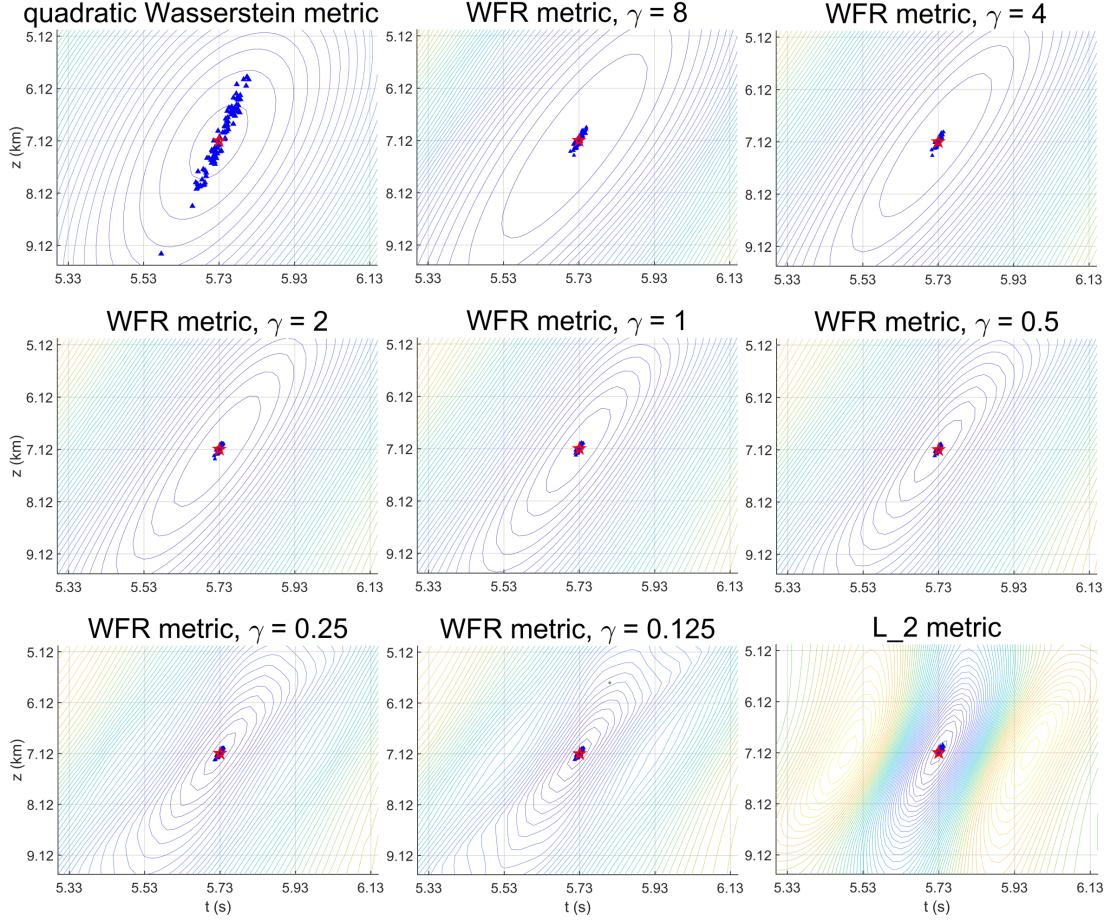


Fig. 3.5. The scattergraph of all the minimum points of the two-layer model with different metrics in Example 3.1 for cross-section at  $x = 46.23$  km of case (i). The blue triangle denotes each minimum point, while the red pentagram denotes the real earthquake hypocenter. The background is a contour plot of the optimization objective function without data noise.

that the WFR metric may offer a better tool for earthquake location problems.

**Remark 3.3.** For the quadratic Wasserstein metric, the accuracy of the earthquake location results of noisy data can be improved by selecting the time window carefully, which is just like we have done in Example 3.1. However, this technique is very sophisticated and requires extra manual operations. In contrast, the WFR metric can achieve more accurate earthquake location results robustly.

**Remark 3.4.** The quadratic Wasserstein metric guarantees global convexity properties. On the contrary, better local convexity holds for the WFR metric, making it less sensitive to the data noise. To fully make use of the advantages of the WFR metric and the classical quadratic Wasserstein metric, we can design a hybrid algorithm. When the initial value and the iterative values are far from the real solution, we can use the classical quadratic Wasserstein metric or the WFR metric with a larger interpolating parameter  $\gamma$ . After iterating for a period of time, the WFR metric with minor interpolating parameter  $\gamma$  is preferred. The comparison in Example 3.1 indicates that the location accuracy is less sensitive to the selection of  $\gamma$  in a certain interval.

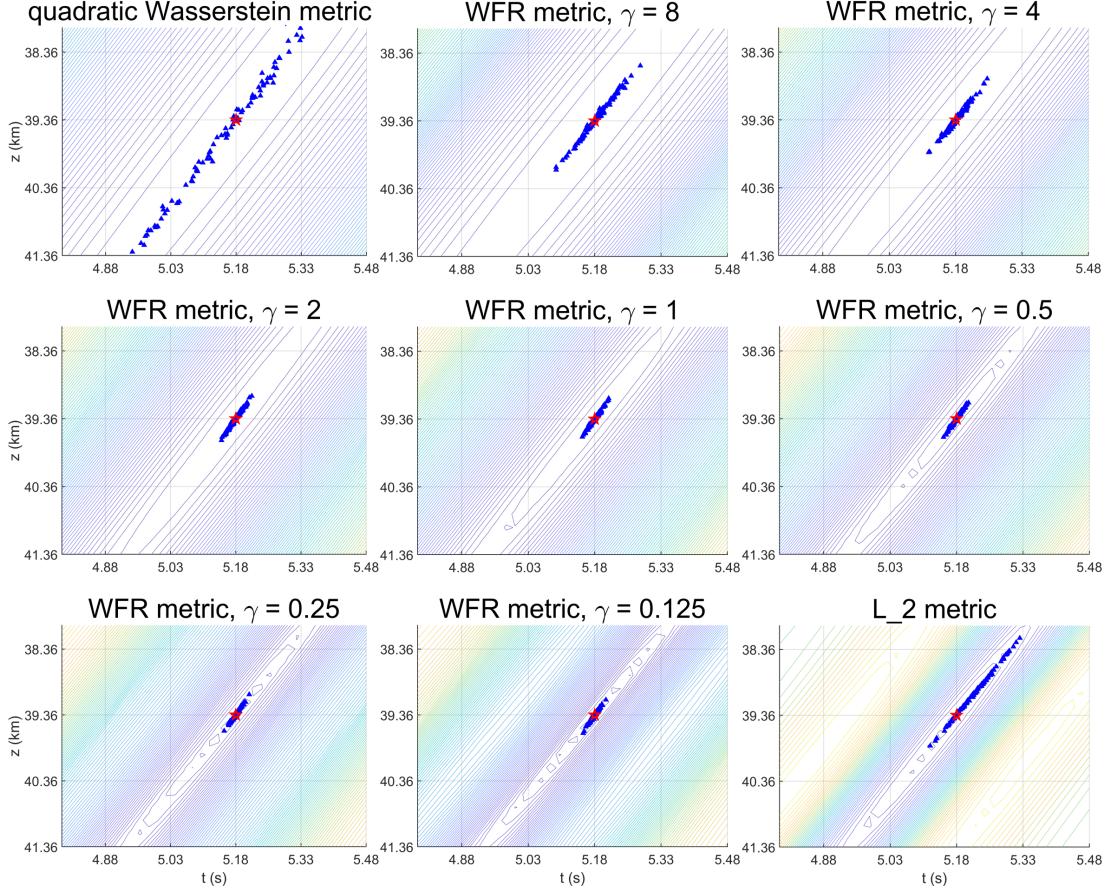


Fig. 3.6. The scattergraph of all the minimum points of the two-layer model with different metrics in Example 3.1 for cross-section at  $x = 57.60$  km of case (ii). The blue triangle denotes each minimum point, while the red pentagram denotes the real earthquake hypocenter. The background is a contour plot of the optimization objective function without data noise.

Thus, in the following numerical experiments in Section 4, the parameter  $\gamma$  is set as 1 at the initial stage of the iteration and then turn to 0.2 at final iterative steps to avoid confusion.

#### 4. Numerical Experiments

In Section 3, the numerical experiments imply higher accuracy of location with the WFR metric when seismic signals are contaminated by noise. In this section, we will use this feature of the WFR metric to the inversion of earthquake hypocenter and origin time. Two numerical experiments are performed to demonstrate the validity and the advantage of the inversion method with the WFR metric.

Firstly, we output the convergence trajectories in the absence of noise. For realistic situations that data is contaminated by noise, the convergence trajectories are output to illustrate the high accuracy of inversion results.

In the numerical examples, the finite difference schemes [10,22] are used to solve the acoustic wave equation (2.4). We apply the perfectly matched layer boundary condition [19] inside the Earth and the reflection boundary condition (2.6) on the Earth's surface. The point source

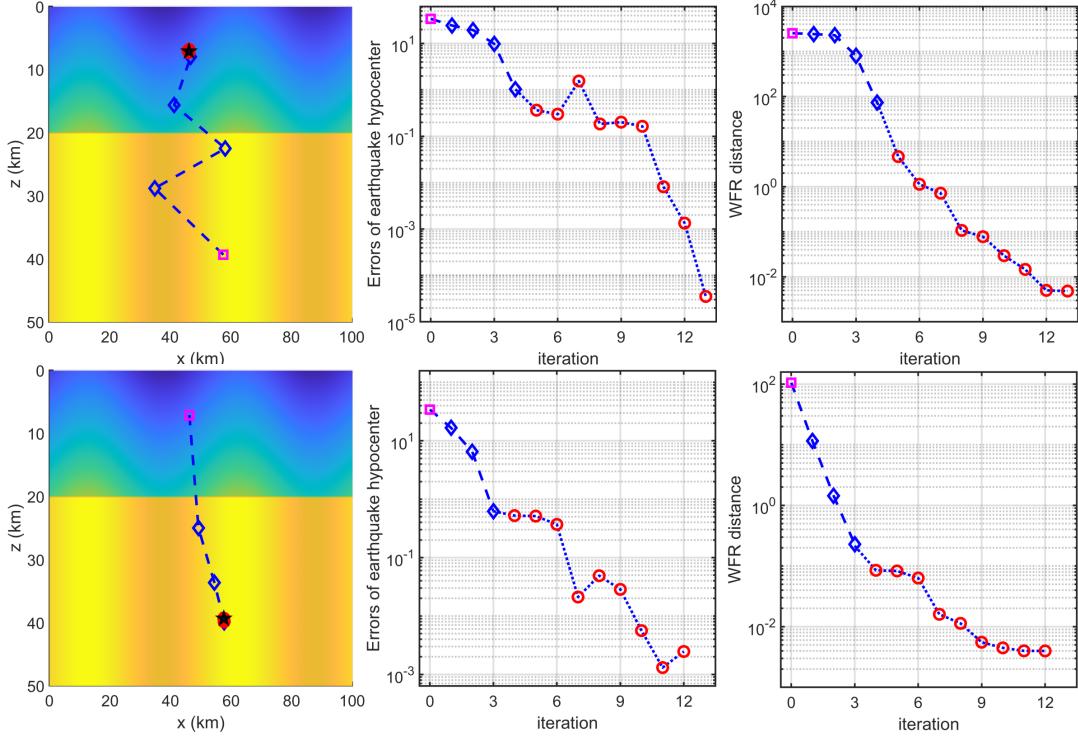


Fig. 4.1. Convergence history of the two-layer model. Up for case (i), and Down for case (ii). Left: the convergent trajectories; Mid: the absolute errors between the real and computed earthquake hypocenter with respect to iteration steps; Right: the WFR distance between the real and synthetic earthquake signals with respect to iteration steps. The magenta square and black pentagram are the initial and the real hypocenter, respectively. The blue diamond and red circle denote the hypocenter in the iterative process with different control parameters  $\gamma = 1$  and  $\gamma = 0.2$ .

$\delta(\mathbf{x} - \boldsymbol{\xi})$  is mathematically singular. Thus we follow the idea from [49] to discretize the delta function by fifth-order piecewise polynomials,

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

Here  $h$  is a numerical parameter related to the mesh size.

#### 4.1. The two-layer model

Consider the same parameters set up as in Example 3.1, the velocity model is illustrated in Figure 3.1. We first study the convergence of this inversion method with the WFR metric under the ideal situation that there is no noise in the data. We consider the case that the earthquake occurs above the Conrad discontinuity, with the inversion starting from the initial hypocenter

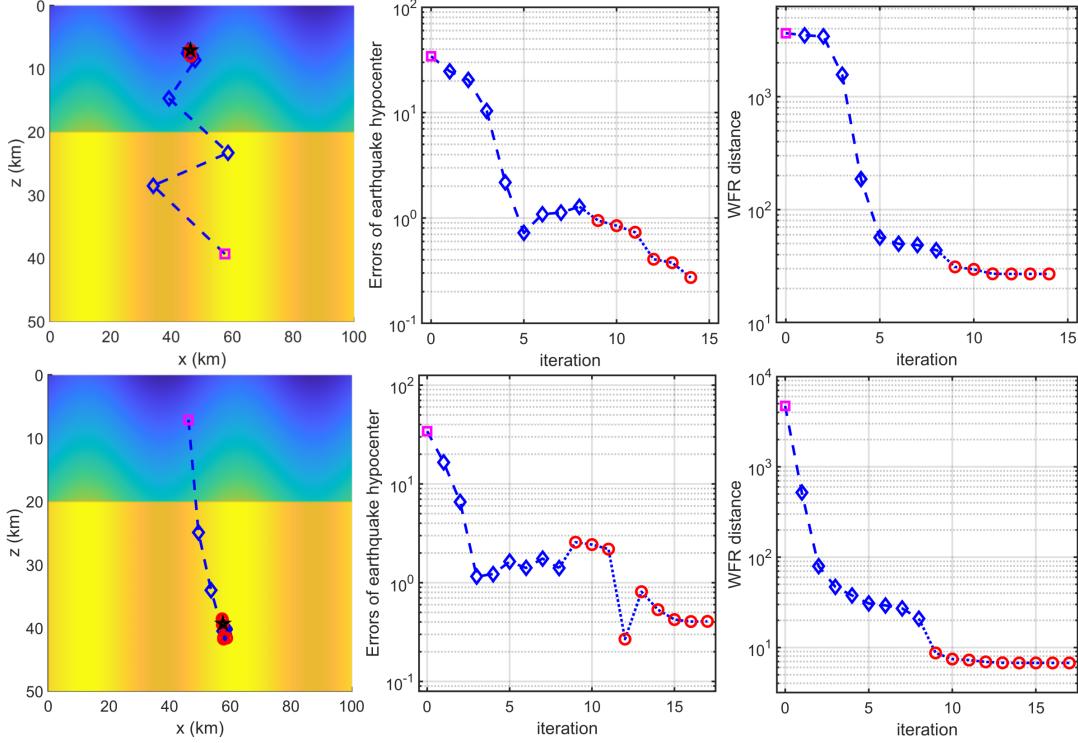


Fig. 4.2. Convergence history of the two-layer model with noise data. Up for case (i), and Down for case (ii). Left: the convergent trajectories; Mid: the absolute errors between the real and computed earthquake hypocenter with respect to iteration steps; Right: the WFR distance between the real and synthetic earthquake signals with respect to iteration steps. The magenta square and black pentagram are the initial and the real hypocenter, respectively. The blue diamond and red circle denote the hypocenter in the iterative process with different control parameters  $\gamma = 1$  and  $\gamma = 0.2$ .

below the Conrad discontinuity (i), and its contrary case (ii).

- (i)  $\xi_T = (46.23 \text{ km}, 7.12 \text{ km})$ ,  $\tau_T = 5.73 \text{ s}$ ,  $\xi = (57.60 \text{ km}, 39.36 \text{ km})$ ,  $\tau = 5.18 \text{ s}$ ,
- (ii)  $\xi_T = (57.60 \text{ km}, 39.36 \text{ km})$ ,  $\tau_T = 5.18 \text{ s}$ ,  $\xi = (46.23 \text{ km}, 7.12 \text{ km})$ ,  $\tau = 5.73 \text{ s}$ .

In Figure 4.1, we can see the convergent trajectories, the absolute errors of the earthquake hypocenter, and the WFR distance. As discussed in Remark 3.4, we select  $\gamma = 1$  to ensure global convexity of the optimization objective function across the interface in the initial stage. When the optimization objective function becomes steady, e.g., decreases less than 10% in one iteration after initial three steps,  $\gamma$  is selected to be 0.2 to ensure better local convexity and non-degeneracy of the optimization objective function near the real earthquake hypocenter and origin time.

Table 4.1: The two-layer model: the location errors of the iterative solution, with unit ‘km’. We compare the results of different cases and different metrics.

	quadratic Wasserstein metric	WFR metric
case (i)	1.28	0.27
case (ii)	1.41	0.41

Table 4.2: The subduction plate model : the horizontal positions of receivers, with unit ‘km’.

$r$	1	2	3	4	5	6
$x_r$	24.53	43.79	47.21	49.96	63.42	97.91
$r$	7	8	9	10	11	12
$x_r$	103.32	120.11	142.92	147.86	166.73	175.35

**Remark 4.1.** In Fig. 4.1, we can observe that the WFR distance always decreases, but the error between the real earthquake hypocenter and the hypocenter computed during the iteration does not keep decreasing. This is because the contour of the WFR distance approximates a very flat ellipse, see Figs. 3.3 and 3.4 for illustration. However, as the iteration step increases, the WFR distance and the error decrease on a relatively long scale.

Next, we investigate the resistance to the noise of the new method. As shown in the equations (3.13)–(3.14) of Example 3.1, real earthquake signals  $d_r(t)$  are obtained by adding white noise  $N_r(t)$  to the noise-free data  $u(\boldsymbol{\eta}_r, t; \boldsymbol{\xi}_T, \tau_T)$ . We select a relatively strong noise  $R = 20\%$ , see Fig. 3.2 for illustration. The convergent histories are output in Fig. 4.2. In Table 4.1, we compare the location errors of different cases and different metrics. Based on the discussions in Example 3.1, we are not surprised that the new earthquake method based on the WFR metric could obtain higher accuracy of the location results when the data is contaminated with noise. Moreover, compared with the noise-free experiments, the iteration numbers remain steady even high-magnitude noise is added to the real signals, which indicates the robustness of this new inversion method.

## 4.2. The subduction plate model

Consider a typical seismogenic zone model discussed in [4, 5, 41, 50]. It consists of the crust, the mantle, and the undulating Moho discontinuity. In addition, there is a subduction zone with a thin low-velocity layer atop a fast velocity layer in the mantle. The earthquakes may occur in any of these areas. Taking into account the complex velocity structure, it is much difficult to locate the earthquake. In the simulating domain  $\Omega = [0, 200 \text{ km}] \times [0, 200 \text{ km}]$ , the wave speed is

$$c(x, z) = \begin{cases} 5.5, & 0 < z \leq 33 + 5 \sin \frac{\pi x}{40}, \\ 7.8, & 33 + 5 \sin \frac{\pi x}{40} < z \leq 45 + 0.4x, \\ 7.488, & 45 + 0.4x < z \leq 60 + 0.4x, \\ 8.268, & 60 + 0.4x < z \leq 85 + 0.4x, \\ 7.8, & \text{others,} \end{cases}$$

with unit ‘km/s’. The simulation mesh size is  $\Delta x = \Delta z = 0.2 \text{ km}$  and time step is  $\Delta t = 0.01 \text{ s}$ . There are 12 randomly distributed receivers  $\boldsymbol{\eta}_r = (\chi_r, z_r)$  on the surface  $z_r = 0 \text{ km}$ . In table 4.2, we output their horizontal positions. This velocity model is illustrated in Fig. 4.3. The dominant frequency of the earthquake is  $f_0 = 2 \text{ Hz}$ , and the amplitude indication parameter is  $A = 15000$ . Their simulating time interval  $I = [0, 55 \text{ s}]$ .

First, consider the ideal situation without noise. We investigate the case when the earthquake occurs in the crust, but the initial guess of the earthquake hypocenter is chosen in the subduction zone. Its contrary case is also taken into account. The parameters are selected as

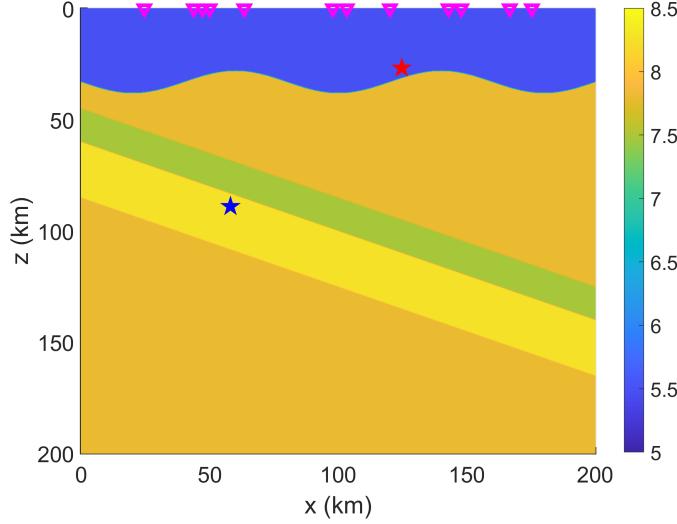


Fig. 4.3. Illustration of the subduction plate model. The inverted magenta triangles indicate the receivers. The red and blue pentagrams indicate the earthquake hypocenter outside and inside the subduction zone, respectively.

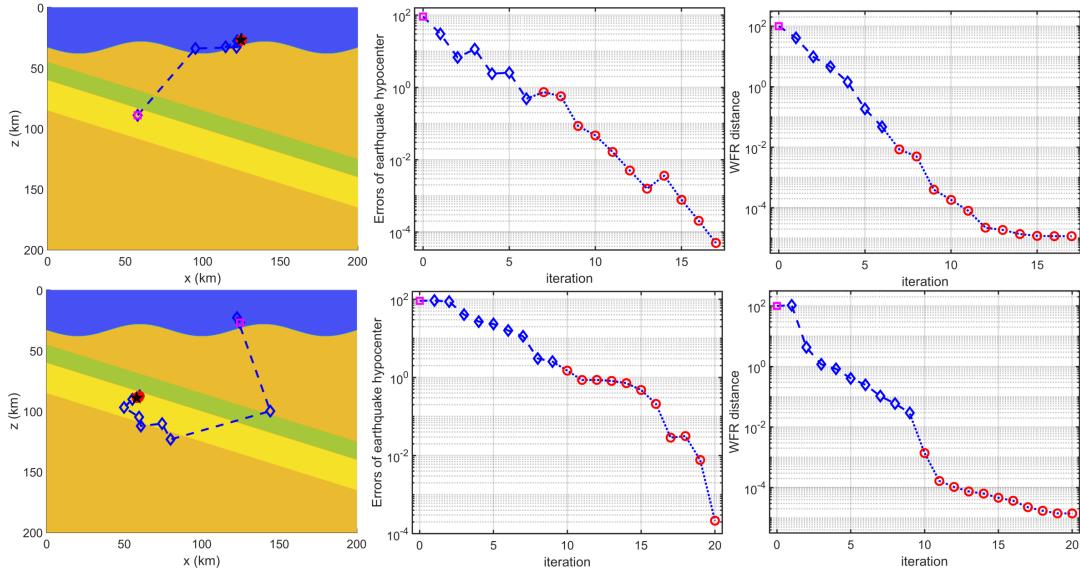


Fig. 4.4. Convergence history of the subduction plate model. Up for case (i), and Down for case (ii). Left: the convergent trajectories; Mid: the absolute errors between the real and computed earthquake hypocenter with respect to iteration steps; Right: the WFR distance between the real and synthetic earthquake signals with respect to iteration steps. The magenta square and black pentagram are the initial and the real hypocenter, respectively. The blue diamond and red circle denote the hypocenter in the iterative process with different control parameters  $\gamma = 1$  and  $\gamma = 0.2$ .

follows:

$$(i) \xi_T = (124.69 \text{ km}, 26.76 \text{ km}), \tau_T = 5.00 \text{ s}, \quad \xi = (58.06 \text{ km}, 88.99 \text{ km}), \tau = 6.79 \text{ s},$$

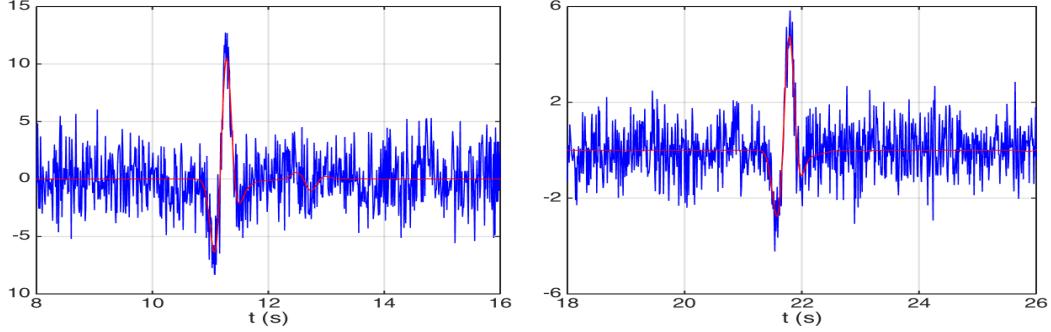


Fig. 4.5. Illustration of signal with noise in the subduction model. The signal with noise  $d_r(t)$  (blue line) and the noise-free signal  $u(\boldsymbol{\eta}_r, t; \boldsymbol{\xi}_T, \tau_T)$  (red line) for receiver  $r = 5$ . The horizontal axis is the time  $t$ . Up: parameters (i) earthquake hypocenter inside the subduction zone; Down: parameters (ii) earthquake hypocenter near the Moho discontinuity.

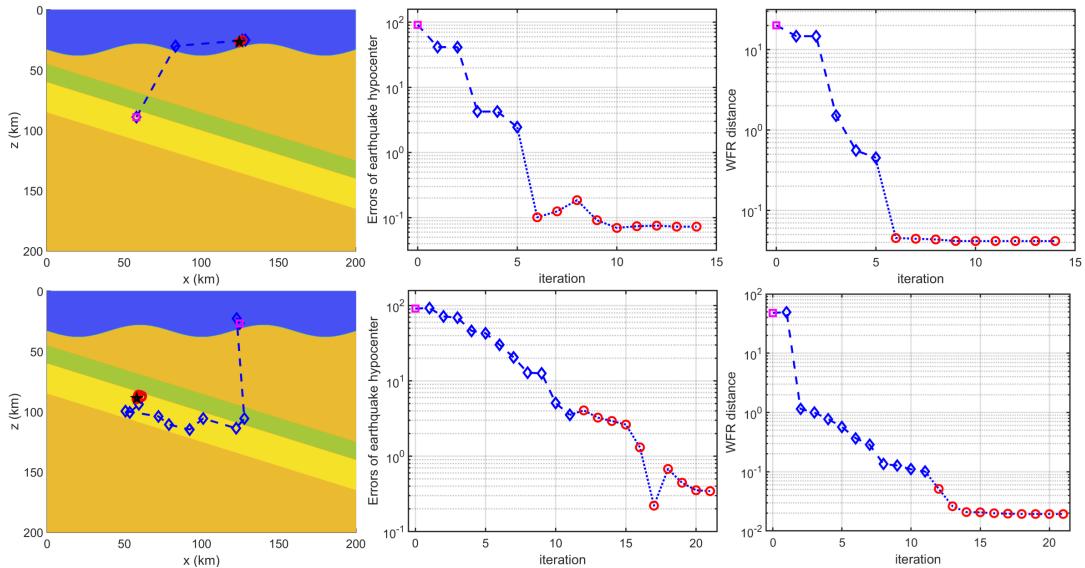


Fig. 4.6. Convergence history of the subduction plate model with noise data. Up for case (i), and Down for case (ii). Left: the convergent trajectories; Mid: the absolute errors between the real and computed earthquake hypocenter with respect to iteration steps; Right: the WFR distance between the real and synthetic earthquake signals with respect to iteration steps. The magenta square and black pentagram are the initial and the real hypocenter, respectively. The blue diamond and red circle denote the hypocenter in the iterative process with different control parameters  $\gamma = 1$  and  $\gamma = 0.2$ .

(ii)  $\boldsymbol{\xi}_T = (58.06 \text{ km}, 88.99 \text{ km})$ ,  $\tau_T = 6.79 \text{ s}$ ,  $\boldsymbol{\xi} = (124.69 \text{ km}, 26.76 \text{ km})$ ,  $\tau = 5.00 \text{ s}$ .

We apply the same  $\gamma$  selection criterion as Section 4.1. The convergent trajectories, absolute errors of the earthquake hypocenter, and the value of WFR distance are output in Fig. 4.4, from which we can observe nice convergence property of the new method.

We consider the same parameters (i) and (ii) but in the situation that signals are contaminated with noise. The noise is added to the real earthquake signals in the same way as in Section 4.1. In Fig. 4.5, we present the real earthquake signal with noise  $d_r(t)$  and the noise-free signal  $u(\boldsymbol{\eta}_r, t; \boldsymbol{\xi}_T, \tau_T)$ , and in Fig. 4.6 we output the convergent history. In Table 4.3, we compare the

Table 4.3: The subduction plate model: the location errors of the iterative solution, with unit ‘km’. We compare the results of different cases and different metrics.

	quadratic Wasserstein metric	WFR metric
case (i)	1.31	0.34
case (ii)	0.54	0.07

location errors of different cases and different metrics. According to this comparison, we can also conclude that the location results of the WFR metric are more accurate than that of the quadratic Wasserstein metric when the data is contaminated with noise.

## 5. Conclusion

What we have seen from the knowledge and experiments above is that the WFR metric is a better choice than the classical quadratic Wasserstein metric for the earthquake location problems. It overcomes the difficulty that the important amplitude information has been ignored by a normalization procedure in the quadratic Wasserstein metric. By introducing the WFR metric, the retained amplitude information provides more constraints for the earthquake location problems. This advantage leads to better convexity property and the non-degeneracy of the optimization objective function near the true source. Thus, this new proposed earthquake location model with the WFR metric can achieve more accurate results when the data is contaminated by strong noise. In the future, we would like to apply this WFR metric to other seismological inverse problems such as velocity inversion and source-velocity joint inversion.

**Acknowledgments.** This work was supported by the National Natural Science Foundation of China (Grant Nos. 11871297, 11971258, U1839206), the National Key Research and Development Program of China on Monitoring, Early Warning and Prevention of Major Natural Disaster (Grant No. 2017YFC1500301), and Tsinghua University Initiative Scientific Research Program. The authors are indebted to Prof. Y. Brenier for his helpful discussions.

## References

- [1] J.D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, *Numer. Math.*, **84** (2000), 375–393.
- [2] J.D. Benamou, G. Carlier, M. Cuturi, L. Nenna and G. Peyré, Iterative Bregman projections for regularized transportation problems, *SIAM J. Sci. Comput.*, **37**:2 (2015), A1111–A1138.
- [3] J.D. Bray, R.B. Seed and H.B. Seed, Analysis of earthquake fault rupture propagation through cohesive soil, *J. Geotech. Engrg.*, **120**:3 (1994), 562–580.
- [4] J. Chen, Y. Chen, H. Wu and D. Yang, The quadratic Wasserstein metric for Earthquake Location, *J. Comput. Phys.*, **373** (2018), 188–209.
- [5] J. Chen, H. Jing, P. Tong, H. Wu and D. Yang, The auxiliary function method for waveform based earthquake location, *J. Comput. Phys.*, **413** (2020), 109453.
- [6] J. Chen, S.K. Kufner, X. Yuan, B. Heit, H. Wu, D. Yang, B. Schurr, and S. Kay, Lithospheric Delamination Beneath the Southern Puna Plateau Resolved by Local Earthquake Tomography, *J. Geophys. Res.: Solid Earth*, **125** (2020), e2019JB019040.
- [7] L. Chizat, G. Peyré, B. Schmitzer and F.X. Vialard, An Interpolating Distance Between Optimal Transport and Fisher-Rao Metrics, *Found. Comput. Math.*, **18**:1 (2018), 1–44.

- [8] L. Chizat, G. Peyré, B. Schmitzer and F.X. Vialard, Scaling Algorithms for unbalanced Optimal Transport Problems, *Math. Comput.*, **87** (2018), 2563–2609.
- [9] F.D. Col, M. Papadopoulou, E. Koivisto, L. Sito, M. Savolainen and L.V. Socco, Application of surface-wave tomography to mineral exploration: a case study from Siilinjärvi, Finland, *Geophys. Prospect.*, **68**:1 (2020), 254–269.
- [10] M.A. Dablain, The application of high-order differencing to the scalar wave equation, *Geophysics*, **51**:1 (1986), 54–66.
- [11] B. Engquist and B.D. Froese, Application of the Wasserstein metric to seismic signals, *Commun. Math. Sci.*, **12**:5 (2014), 979–988.
- [12] B. Engquist, B.D. Froese and Y. Yang, Optimal transport for seismic full waveform inversion, *Commun. Math. Sci.*, **14**:8 (2016), 2309–2330.
- [13] B. Engquist, K. Ren and Y. Yang, The quadratic Wasserstein metric for inverse data matching, *Inverse Probl.*, **36**:5 (2020), 055001.
- [14] B. Engquist and Y. Yang, Seismic imaging and optimal transport, *Commun. Inf. Syst.*, **19**:2 (2019), 95–145.
- [15] T.O. Gallouët, M. Laborde and L. Monsaingeon, An unbalanced Optimal Transport splitting scheme for general advection-reaction-diffusion problems, *ESAIM: COCV*, **25** (2019), 8.
- [16] W. Gangbo, W. Li, S. Osher and M. Puthawala, Unnormalized optimal transport, *J. Comput. Phys.*, **399** (2019), 108940.
- [17] M.C. Ge, Analysis of source location algorithms Part I: Overview and non-iterative methods, *J. Acoust. Emiss.*, **21** (2003), 14–28.
- [18] S. Kondratyev, L. Monsaingeon and D. Vorotnikov, A new optimal transport distance on the space of finite Radon measures, *Adv. Differ. Equat.*, **21** (2016), 1117–1164.
- [19] D. Komatitsch and J. Tromp, A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation, *Geophys. J. Int.*, **154** (2003), 146–153.
- [20] J.J. Kosowsky and A.L. Yuille, The invisible hand algorithm: Solving the assignment problem with statistical physics, *Neural Netw.*, **7**:3 (1994), 477–490.
- [21] T. Le, M. Yamada, K. Fukumizu and M. Cuturi, Tree-Sliced Variants of Wasserstein Distances, *arXiv:1902.00342v3*, 2019.
- [22] J. Li, D. 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** (2017), 1938–1964.
- [23] R. Li and F. Yang, A reconstructed discontinuous approximation to Monge-Ampere equation in least square formation, *arXiv:2010.09921v3*, 2019.
- [24] M. Liero, A. Mielke and G. Savaré, Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves, *SIAM J. Math. Analysis*, **48**:4 (2016), 2869–2911.
- [25] M. Liero, A. Mielke and G. Savaré, Optimal Entropy-Transport problems and a new Hellinger-Kantorovich distance between positive measures, *Invent. Math.*, **211** (2018) 969–1117.
- [26] Q. Liu, J. Polet, D. Komatitsch and J. Tromp, Spectral-Element Moment Tensor Inversion for Earthquakes in Southern California, *Bull. seism. Soc. Am.*, **94**:5 (2004), 1748–1761.
- [27] R. Madariaga, Seismic Source Theory, in *Treatise on Geophysics (Second Edition)*, S. Gerald (ed.), Elsevier B.V., 2015, 51–71.
- [28] C. Meng, Y. Ke, J. Zhang, M. Zhang, W. Zhong and P. Ma, Large-scale optimal transport map estimation using projection pursuit , *Adv. Neural Inf. Process. Syst.*, **32** (2019), 8118–8129.
- [29] C. Meng, J. Yu, J. Zhang, P. Ma and W. Zhong, Sufficient dimension reduction for classification using principal optimal transport direction, *arXiv:2010.09921v3*, 2020.
- [30] 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** (2016), 345–377.

- [31] 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 Probl.*, **32** (2016), 115008.
- [32] W. Pan and Y. Wang, On the influence of different misfit functions for attenuation estimation in viscoelastic full-waveform inversion: synthetic study, *Geophys. J. Int.*, **221**:2 (2020), 1292–1319.
- [33] G. Peyré and M. Cuturi, Computational Optimal Transport: With Applications to Data Science, *Found. Trends Mach. Learn.*, **11**:5–6 (2019), 355–607.
- [34] B. Piccoli and F. Rossi, Generalized Wasserstein Distance and its Application to Transport Equations with Source, *Arch. Rational Mech. Anal.*, **211** (2014), 335–358.
- [35] L. Qiu, J. Ramos-Martinez, A. Valenciano, Y. Yang and B. Engquist, Full-waveform inversion with an exponentially encoded optimal-transport norm, *SEG Tech. Program Expanded Abstr.*, (2017) 1286–1290.
- [36] N. Rawlinson, S. Pozgay and S. Fishwick, Seismic tomography: A window into deep Earth, *Phys. Earth Planet. Inter.*, **178** (2010), 101–135.
- [37] F. Santambrogio, Optimal Transport for Applied Mathematicians, Birkhäuser, 2015.
- [38] C. Satriano, A. Lomax and A. Zollo, Real-Time Evolutionary Earthquake Location for Seismic Early Warning, *Bull. seism. Soc. Am.*, **98**:3 (2008), 1482–1494.
- [39] B. Schmitzer, Stabilized Sparse Scaling Algorithms for Entropy Regularized Transport Problems, *SIAM J. Sci. Comput.*, **41**:3 (2019), A1443–A1481.
- [40] M. Sharify, S. Gaubert, and L. Grigori, Solution of the optimal assignment problem by diagonal scaling algorithms, *arXiv:1104.3830*, 2013.
- [41] P. Tong, D. Zhao and D. Yang, Tomography of the 1995 Kobe earthquake area: comparison of finite-frequency and ray approaches, *Geophys. J. Int.*, **187** (2011), 278–302.
- [42] P. Tong, D. Zhao, D. Yang, X. Yang, J. Chen and Q. Liu, Wave-equation-based travel-time seismic tomography - Part 1: Method, *Solid Earth*, **5** (2014), 1151–1168.
- [43] P. Tong, D. Zhao, D. Yang, X. Yang, J. Chen and Q. Liu, Wave-equation-based travel-time seismic tomography C Part 2: Application to the 1992 Landers earthquake (Mw 7.3) area, *Solid Earth*, **5** (2014), 1169–1188.
- [44] P. Tong, D. Yang, Q. Liu, X. Yang and J. Harris, Acoustic wave-equation-based earthquake location, *Geophys. J. Int.*, **205** (2016), 464–478.
- [45] C. Villani, *Optimal Transport: Old and New*, Springer Science & Business Media, 2008.
- [46] F. Waldhauser and W.L. Ellsworth, A double-difference earthquake location algorithm: Method and application to the northern Hayward Fault, California, *Bull. seism. Soc. Am.*, **90**:6 (2000), 1353–1368.
- [47] X.J. Wang, On the design of a reflector antenna II, *Calc. Var. Partial Dif.*, **20**:3 (2004), 329–341.
- [48] Z. Wang, D. Zhou, M. Yang, Y. Zhang, C. Bao and H. Wu, Robust Document Distance with Wasserstein-Fisher-Rao Metric, in *Proceedings of The 12th Asian Conference on Machine Learning*, PMLR 129, 2020, 721–736.
- [49] X. Wen, High Order Numerical Quadratures to One Dimensional Delta Function Integrals, *SIAM J. Sci. Comput.*, **30**:4 (2008), 1825–1846.
- [50] H. Wu, J. Chen, X. Huang and D. Yang, A new earthquake location method based on the waveform inversion, *Commun. Comput. Phys.*, **23**:1 (2018), 118–141.
- [51] H. Wu and X. Yang, Eulerian Gaussian beam method for high frequency wave propagation in the reduced momentum space, *Wave Motion*, **50**:6 (2013), 1036–1049.
- [52] Y. Yang, B. Engquist, J. Sun and B.D. Froese, Application of Optimal transport and the quadratic Wasserstein metric to Full-Waveform-Inversion, *Geophysics*, **83**:1 (2018), R43–R62.
- [53] Y. Yang and B. Engquist, Analysis of optimal transport and related misfit functions in full-waveform inversion, *Geophysics*, **83**:1 (2018), A7–A12.
- [54] Z. Yan and Y. Wang, Full waveform inversion with sparse structure constrained regularization, *J.*

- Inverse Ill-Posed Probl.*, **26**:2 (2018), 243–257.
- [55] X. Zhao, Z. Wang, Y. Zhang and H. Wu A Relaxed Matching Procedure for Unsupervised BLI, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2020, 3036–3041.
  - [56] W. Zhang, Acoustic multi-parameter full waveform inversion based on the wavelet method, *Inverse Probl. Sci. Eng.*, **29**:2 (2021), 220–247.
  - [57] W. Zhang and J. Luo, Full-waveform velocity inversion based on the acoustic wave equation, *Am. J. Comput. Math.*, **3** (2013), 13–20.