Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2024-0230

Vol. **40**, No. 1, pp. 129-152
July 2026

Full-Waveform Inversion with Unbalanced Optimal Transport Metric for Seismic Imaging

Liangrong Wang[1], Xingpeng Dong[2] and Dinghui Yang[1,*]

[1] *Department of Mathematical Sciences, Tsinghua University, Beijing 100084, P.R. China.*
[2] *Institute of Geology, China Earthquake Administration, Beijing 100029, P.R. China.*

Received 27 September 2024; Accepted (in revised version) 30 March 2025

Abstract. Full-waveform inversion (FWI) is an effective method for obtaining high-resolution images of subsurface structures. Conventional full-waveform inversion, which uses the least-square norm L_2 to measure the mismatch between observed and synthetic seismograms, frequently suffers from cycle-skipping and local minimum problems. Derived from optimal transport theory, the Wasserstein metric has been proposed to mitigate cycle-skipping issue. However, due to the requirement for mass conservation, the classical quadratic Wasserstein metric is not ideally suited for FWI applications. In this study, we introduce two unbalanced optimal transport (UOT) distances for use in FWI: the regularized UOT and the unbalanced Sinkhorn divergence. An entropy regularization approach and a truncation approximation are employed to guarantee the efficiency of calculating distances and gradients. These unbalanced optimal transport distances preserve the desirable properties of the quadratic Wasserstein metric, particularly its convexity and insensitivity to noise, while overcoming issues related to mass conservation. We compare the unbalanced optimal transport distances with the L_2 distance and the classical quadratic Wasserstein metric using the Camembert model and the crustal root model. Our numerical experiments demonstrate the superiority of the unbalanced optimal transport distances over traditional methods.

AMS subject classifications: 65K10, 86-08, 86A15, 86A22

Key words: Full-waveform inversion, Wasserstein metric, unbalanced optimal transport, Sinkhorn divergence.

1 Introduction

Full-waveform inversion (FWI), first proposed by Lailly [21] and Tarantola [41], is a data fitting procedure to obtain high-resolution information of subsurface structure. Different

---

[*] Corresponding author. *Email addresses:* `ydh@tsinghua.edu.cn` (D. Yang), `wlr20@mails.tsinghua.edu.cn` (L. Wang), `dongxp@ies.ac.cn` (X. Dong)

from ray-based methods [1], FWI estimates subsurface properties by solving the wave equations, thus it can acquire comprehensive information of seismic waves. It is now widely used in seismic tomography [3,24,32,44,47,51].

Mathematically, FWI is a nonlinear optimization problem constrained with partial differential equations. The model is iteratively updated by minimizing the misfit function of synthetic data and observed data. Because of the high-dimensionality of model space, FWI is typically solved using gradient-based optimization methods in practice. The gradient of the misfit function is updated using the adjoint-state method [34,42]. In conventional FWI, the least-square objective function L_2 is used for measuring the data misfit [5,14,40,44,47]. However, the inverse problem is ill-posed and the L_2 norm is only a point-by-point measurement of the amplitude difference. When the initial model is far from the true model, i.e., the travel time error between synthetic and observed data is more than a half period, an incorrect velocity model will be generated [31]. In other words, the optimization falls into a local minimum. This phenomenon is called cycle-skipping [15,44], resulting in not to converge to the global minimum. Therefore, FWI employing the L_2 distance requires an accurate enough initial model, which is generally impossible in actual situations.

In the last decades, several approaches have been developed to modify the misfit function to overcome the defect of the L_2 distance. Luo and Schuster [27] first proposed to use the cross-correlation function of synthesis and observations to estimate the traveltime residual as the misfit function. This method is more robust and widely utilized in finite-frequency tomography [42]. However, the cross-correlation measurement requires that the signals have approximately the same shape. To address this issue, a misfit function based on deconvolution has been developed [26,46]. In this approach, the traveltime misfit is computed by deconvolving the synthetic data with the observed data, rather than relying on cross-correlation. Another strategy is the envelope misfit [25,50], which measures the instantaneous phase and amplitude envelope using the Hilbert transform. More recently, Dong and Yang [8] have redefined the traditional L_2 norm by incorporating a time shift determined by cross-correlation within the synthetic waveform, resulting in phase-sensitive full-waveform tomography. These strategies aim to improve the posedness of the FWI problem by constructing more convex misfit functions. However, these misfit functions essentially measure the L_2 distance between preprocessed data. Although these methods relax in some manner the dependency on the accuracy of the initial model, cycle-skipping may still occur [28]. Recently, Engquist and Froese [11] first applied optimal transport (OT) mapping to seismic tomography. The Wasserstein metric, derived from OT [43], is defined as the misfit function for FWI. The optimal transport problem was proposed by Monge in 1781, in order to search for the optimal way of transporting sand. Kantorovitch [16] relaxes the original nonlinear problem into a linear optimization problem with convex constraints. The objective function of the problem is known as the Wasserstein metric. Because OT-based techniques can incorporate differences in spatial information, OT-based measures have recently gained widespread use in various applications, including image retrieval [23], signal and image