# Optimal Transport for Elastic Seismic Source Inversion

**Tyler Masthay**[1,*], **Björn Engquist**[1]
[1]Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, USA
*Email: tyler@oden.utexas.edu

## Abstract

Full-waveform inversion (FWI) is a state-of-the-art method for imaging the earth's subsurface. However, FWI is notorious for local-minimum trapping, or "cycle skipping," and thus requires an accurate initial model ([2]). Cycle skipping is caused by the nonconvex nature of the misfit optimization landscape in its typical least-squares formulation. The Wasserstein-2 distance is convex with respect to shifts and dilations, both of which occur naturally in seismic data. Therefore, we propose using this optimal transport metric as our misfit for FWI. Previous work using optimal transport for source inversion, whose applications include microseismic event detection and deformation mechanics in subduction zones, has shown promise ([1]). However, this work uses the acoustic wave equation, which is less accurate than the elastic wave equation. In this paper, we extend these results to elastic source inversion and show that they translate well to the elastic model.

**Keywords: seismic imaging, optimal transport, inverse problems**

## 1 Introduction

FWI is formulated as a PDE-constrained optimization problem with respect to a given misfit functional. Given that time shifts and amplitude dilations occur naturally in seismic data, we would ideally use a misfit functional that is convex with respect to these transformations. The Wasserstein-2 distance, denoted $W_2$, satisfies this property, whereas the industry standard $L^2$ misfit does not ([4]). $W_2(\mu, \nu)$ for probability distributions $\mu, \nu$ on $X = \mathbb{R}^n$ is given by

$$W_2^2(\mu, \nu) := \inf_{T \in \mathcal{M}} \int_X |x - T(x)|^2 d\mu(x) \qquad (1)$$

where

$$\mathcal{M} := \{T | \forall B \in \mathcal{B}(X), \mu(T^{-1}(B)) = \nu(B)\}$$

is the set of feasible transport maps ([3]). $W_2$ is defined between probability distributions, but

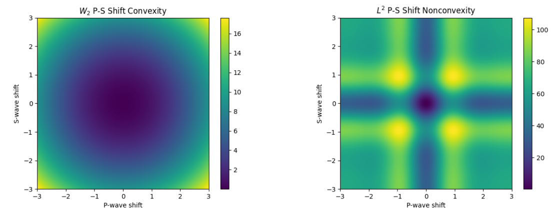


Figure 1: Comparison of optimization landscape of $W_2$ and $L^2$ distances for elastic waves that have been shifted away from origin according to varying $P$ and $S$-wave velocity differences.

seismic data are not probability distributions. Thus, we must renormalize our seismic data via some map $\mathcal{R}$; we use positive/negative splitting renormalization as outlined in [4]. Figure 1 demonstrates that convexity of $W_2$ under shift is preserved after renormalization of a model elastic wave with a Ricker wavelet source. Directly computing the $L^2$ norm clearly results in a nonconvex optimization landscape. Previous work has applied optimal transport to velocity and source inversion but has used the acoustic wave equation as the forward model ([1], [5]). In this paper, we apply optimal transport to seismic source inversion but with the elastic wave equation, a more complete and accurate forward model. We exhibit that the promising results from the acoustic model translate well to the elastic model.

## 2 Problem Formulation

Our forward model is the isotropic elastic wave equation over domain $\Omega = (a, b) \times (c, d) \times (0, T)$. Our Lamé parameters are given by $\lambda(\boldsymbol{x})$ and $\mu(\boldsymbol{x})$. The source signature of a Ricker wavelet with a characteristic timescale $\sigma$ and amplitude $A$. The model parameter is the source location $\boldsymbol{s}$. That is, $\boldsymbol{u} = \mathcal{F}(\boldsymbol{s})$ if and only if

$$\begin{aligned} \rho \ddot{\boldsymbol{u}} &= (\lambda + \mu)\nabla(\nabla \cdot \boldsymbol{u}) + \mu \nabla^2 \boldsymbol{u} + \boldsymbol{f}_{\boldsymbol{s}} \\ &\quad + \nabla\lambda(\nabla \cdot \boldsymbol{u}) + \nabla\mu \cdot (\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T) \text{ in } \Omega \\ u(z, x, 0) &= 0 \\ \boldsymbol{u} & \text{ satisfies Sommerfeld condition} \end{aligned}$$

where

$$\boldsymbol{f_s} = AR_\sigma(t)\boldsymbol{\delta}(\boldsymbol{x}-\boldsymbol{s})$$
$$R_\sigma(t) := \left(1-\left(\frac{t}{\sigma}\right)^2\right)e^{-\frac{t^2}{2\sigma^2}}.$$

Given that $W_2$ computation has linear cost in 1D and is expensive in higher dimensions, we use the trace-by-trace $W_2$ distance, as outlined in [4]. The trace-by-trace $W_2$ distance computes the renormalized $W_2$ distance between time series at each receiver location and then sums them. Formally, we define an observation map $\mathcal{O}$ by

$$\mathcal{O}(\boldsymbol{u})(x) := \boldsymbol{u}(0,x,\cdot). \tag{2}$$

Thus our full map from parameter space to observation space is given by

$$\mathcal{P}_{\boldsymbol{s}} := \mathcal{R}\left[\mathcal{O}\left[\mathcal{F}(\boldsymbol{s})\right]\right]. \tag{3}$$

Given observation time series $\{\boldsymbol{d}_r(t) : t \in (0,T)\}_{r=1}^R$ at surface coordinates $\{x_r\}_{r=1}^R$, we seek a source location $\boldsymbol{s}^*$ such that

$$\boldsymbol{s}^* = \arg\min_{\boldsymbol{s}} \sum_{r=1}^R W_2^2\left(\mathcal{P}_{\boldsymbol{s}}(x_r), \mathcal{R}(\boldsymbol{d}_r)\right).$$

## 3 Computational Results

In Figure 3, we compare the optimization landscape of our modified $W_2$ distance and the $L^2$ norm. Note that the $W_2^2$ landspace is much smoother and has a unique global minimum. This contrasts to the many spiky local minima seen for the square of the $L^2$ norm. Given an inaccurate initial source location, we would expect convergence to the global minimum for $W_2^2$, but we will likely see convergence to only a local minimum for $L^2$. In Figure 3, we directly test this with an initial guess of $(0.8, 0.5)$ with data $\boldsymbol{d} = \mathcal{F}((0.5, 0.5))$ synthetically generated at the center of the domain. We see convergence to the global minimum for $W_2$ in few iterations and only convergence to a local minimum (that is quite far away from the global minimum) for $L^2$. Our experiments support our hypothesis that when applied to the elastic wave equation, $W_2$ has a smoother optimization landscape with less likelihood of getting trapped in a local minimum.

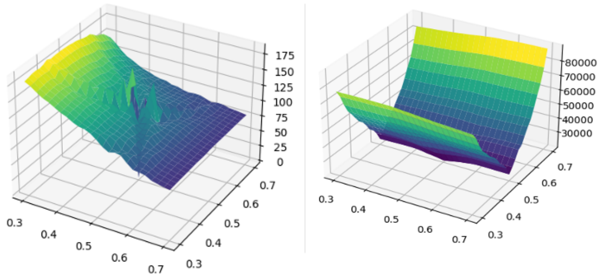

Figure 2: (left) $L^2$ squared distance v. source location, (right) $W_2^2$ v. source location. Synthetic data generated from source location at center of domain $(0.5, 0.5)$, where global minimum is located.

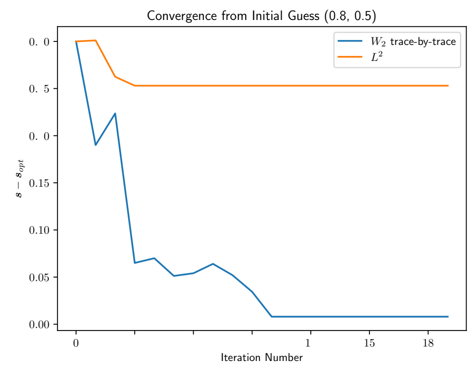


Figure 3: Convergence to global minimum for $W_2$ versus convergence to local minimum for $L^2$

## References

[1] Jing Chen, Yifan Chen, Hao Wu, and Dinghui Yang. The quadratic wasserstein metric for earthquake location. *Journal of Computational Physics*, 373:188–209, 2018.

[2] Ludovic Métivier, Aude Allain, Romain Brossier, Quentin Mérigot, Edouard Oudet, and Jean Virieux. Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach. *Geophysics*, 83(5):R515–R540, 2018.

[3] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

[4] Yunan Yang and Björn Engquist. Analysis of optimal transport and related misfit functions in full-waveform inversion. *Geophysics*, 83(1):A7–A12, 2018.

[5] Yunan Yang, Björn Engquist, Junzhe Sun, and Brittany F Hamfeldt. Application of optimal transport and the quadratic wasserstein metric to full-waveform inversion. *Geophysics*, 83(1):R43–R62, 2018.