Full waveform inversion with unbalanced optimal transport distance

Da Li, Michael P. Lamoureux and Wenyuan Liao

Department of Mathematics and Statistics, University of Calgary

Summary

Full waveform inversion (FWI) has become a major seismic imaging technique. However, using the least-squares norm in the misfit functional possibly leads to cycle-skipping issue and increases the nonlinearity of the optimization problem. Several works of applying optimal transport distances to mitigate this problem have been proposed recently. The optimal transport distance is to compare two positive measures with equal mass. To overcome the mass equality limit, we introduce an unbalanced optimal transport (UOT) distance with Kullback–Leibler divergence to balance the mass difference. An entropy regularization and a scaling algorithm have been used to compute the distance and its gradient efficiently. Numerical examples are provided to show the behavior of UOT distance applied in FWI problem and comparison with L2 distance has been provided.

Introduction

Full waveform inversion (FWI) is a high-resolution seismic imaging algorithm and it was proposed in the early 1980s. It is a nonlinear PDE-constrained optimization problem with physical properties such as velocity and density of underground as the control parameters, and the waveform received by the receivers as the state parameters. Depending on different physical model, the constraint PDE can be simple wave equation, acoustic wave equation or elastic wave equation. Because of the huge size of the scale, gradient based optimization methods such as gradient descent, l-BFGS and Newton method is needed. And the gradient generally can be achieved by the adjoint state method. With the improvement of the computing power, FWI has been more and more applied in the industry.

In conventional methods, the L2 distance is used in the misfit function during optimization to measure the difference between observed and synthetic data. As a nonlinear optimization problem, FWI algorithm suffers the existence of local minima. One of the reasons causing the local minima is cycle-skipping issue, which can occur as the phase difference between two seismic signals is larger than half wavelength. To mitigate this problem, using optimal transport (OT) distances (Wasserstein distance) in FWI problem have been proposed recently. The optimal transport distance is to compare two positive measures with equal total mass. When comparing two non-negative equal mass functions, the OT distance will keep monotonically increasing as one function is shifting away from another function. This property provides convexity of OT distance as a misfit functional and it is one of the main reasons to introduce OT distance to FWI problem (Engquist et al., 2016). Several works have been proposed of applying OT distance to FWI problem recently (Métivier et al., 2016a; Yang et al., 2018; Yong et al., 2019).

In this work we consider the FWI problem with wave equation as the constraint. We introduce the unbalanced optimal transport (UOT) distance to remove the equal mass restriction. To compute the distance and gradient efficiently, an entropy regularization method and a scaling method have been used.

Theory

1. Unbalanced Optimal Transport Problem

For two discrete vectors f and g, the optimal transport problem is defined as:

$$W_2^2(f,g) = \min_{T\in\mathbb{R}^{N\times N}} < T, C > = \sum_{i,j=1}^{N} T_{i,j}C_{i,j}\,, \qquad s.t. T1_N = f, \qquad T^T 1_N = g. \qquad (1)$$

Here the cost matrix C is defined as $C_{i,j} = |x_i - x_j|^2$, and the $W_2(f,g)$ is called 2-Wasserstein distance. One of the disadvantages of optimal transport is nonnegative measures with the same total mass are required. To overcome this limitation, we introduce the unbalanced optimal transport (UOT) distance mainly based on the work in (Chizat et al., 2018). The UOT distance with entropy regularization can be represented as:

$$W_{2,\varepsilon,\varepsilon_m}^2(f,g) = \min_{T\in\mathbb{R}^{N\times N}} < T, C > -\, \varepsilon E(T) + \varepsilon_m KL(T1_N|f) + \varepsilon_m KL(T^T 1_N|g). \qquad (2)$$

Here $E(T) = -\sum_{i,j} T_{i,j}(\log T_{i,j} - 1)$ is the entropy regularization for computational efficiency. And $KL(f|g) = \sum_i f_i(\log f_i/g_i - 1)$ is the Kullback-Leibler divergence which measures the difference between vector or matrix f and g. Equation (2) is equivalence to:

$$W_{2,\varepsilon,\varepsilon_m}^2(f,g) = \min_{T\in\mathbb{R}^{N\times N}} \varepsilon KL(T|K) + +\varepsilon_m KL(T1_N|f) + \varepsilon_m KL(T^T 1_N|g). \qquad (3)$$

Where $K_{i,j} = \exp\left(-\frac{C_{i,j}}{\varepsilon}\right)$. The dual formula of equation (3) is given by:

$$W_{2,\varepsilon,\varepsilon_m}^2(f,g) = \max_{\phi,\psi\in\mathbb{R}_+^N} \sum_{i,j=1}^{N} -\varepsilon_m f_i\left(e^{-\frac{\phi_i}{\varepsilon_m}-1}\right) - \varepsilon_m g_i\left(e^{-\frac{\psi_j}{\varepsilon_m}-1}\right) - \varepsilon K\left(e^{\frac{\phi_i}{\varepsilon}} e^{\frac{\psi_j}{\varepsilon}} - 1\right). \qquad (4)$$

Also, if ϕ^* and ψ^* is the solution of equation (4), then:

$$T_{i,j}^* = e^{\frac{\phi_i^*}{\varepsilon}} K_{i,j} e^{\frac{\psi_j^*}{\varepsilon}}.$$

To compute the regularized UOT distance, a coordinate ascent method can be used. Starting from the initial value $v^{(0)} = 1_N$, the dual problem (4) can be computed with iterating:

$$u_i^{(n+1)} = \left(\frac{f_i}{\sum_j K_{i,j} v_j^{(n)}}\right)^{\frac{\varepsilon_m}{\varepsilon_m+\varepsilon}}, \qquad v_j^{(n+1)} = \left(\frac{g_j}{\sum_i K_{i,j} u_i^{(n)}}\right)^{\frac{\varepsilon_m}{\varepsilon_m+\varepsilon}}.$$

Here $u_i^{(n)} = e^{\frac{\phi_i^n}{\varepsilon}}$ and $v_j^{(n)} = e^{\frac{\psi_j^n}{\varepsilon}}$. As soon as the solution is found, the gradient is given by:

$$\nabla_{f_i} W_{2,\varepsilon,\varepsilon_m}^2(f,g) = -\varepsilon_m\left(e^{-\frac{\phi_i^*}{\varepsilon_m}-1}\right). \qquad (5)$$

2. Full Waveform Inversion

Since optimal transport problem was proposed for positive measures, exponential normalization is used in this paper:

$$h(f) = e^{kf}.$$

Here k is the normalization coefficient. The full waveform inversion problem can be described as a PDE constraint optimization problem:

$$\min_{c} J(c) = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} W_{2,\varepsilon,\varepsilon_m}^2 \left(h(P_r u_s), h\left(P_r d_{obs,s}\right) \right).$$

Here P_r is the projection operator maps the wavefield to the rth receiver to record seismic data. The objective function is constrained by N_s wave equations:

$$\frac{1}{c^2}(u_s)_{tt} - \Delta \mathrm{u_s} = f_s, \;\; s = 1, \dots, N_s.$$

To achieve the gradient of the objective function, adjoint equations need to be solved:

$$\frac{1}{c^2}(v_s)_{tt} - \Delta \mathrm{v_s} = -\sum_{r=1}^{N_r} P_r^T \left(k e^{k P_r u_s}\right)^T \nabla W_{2,\varepsilon,\varepsilon_m}^2 \left(h(P_r u_s), h\left(P_r d_{obs,s}\right) \right).$$

The gradient of UOT distance is given by equation (5). Then the gradient of objective function is:

$$\nabla J(c) = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} -\frac{2}{c^3} (u_s)_{tt} v_s \,.$$

Numerical Results

1. One dimensional example

This example shows the UOT distance is more sensitive with respect to time shift of signal than the conventional L2 distance. Figure 1 plots two Ricker wavelets f and g. The signal g is fixed at $t = 0.5$ and we compute the L2 and UOT distance between f and g when f is shifting from $t = 0.3$ to $t = 0.7$. The L2 and UOT difference are shown in Figure 2 (a) and (c). With proper normalization coefficient, the UOT difference can has a monotone behavior as the distance between f and g increases. Figure 2 (b) and (d) show the adjoint source with L2 and UOT distance. Comparing to L2 adjoint source, the energy of UOT adjoint sources concentrate on the position of the wavelet and providing less detail. This effect will lead to a smoother gradient during the optimization.

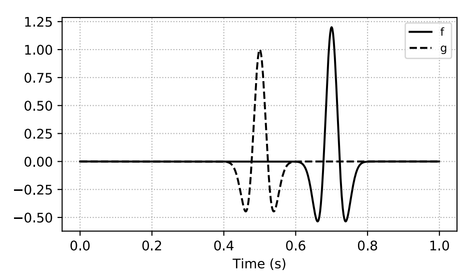

Figure 1. Two Ricker wavelets.

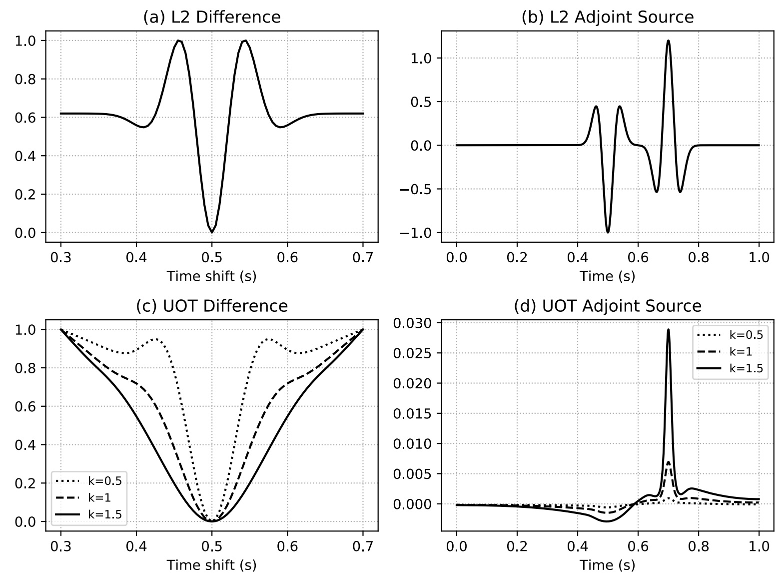

Figure 2. (a): L2 difference. (b): L2 adjoint source. (c): UOT difference with different normalization coefficients. (d) UOT adjoint sources with difference normalization coefficients.

2. Marmousi 2 Model

We perform FWI with L2 and UOT distance on Marmousi 2 data. The true model and initial model are shown in Figure 3. There are 11 equally spaced sources and 101 equally spaced receivers located on the top of the model. A 5 Hz Ricker wavelet is used as the source function. The nonlinear conjugate gradient is used during the optimization. Figure 4 shows the L2 and UOT gradient at the first optimization step. Comparing to the L2 gradient, UOT gradient provides more large-scale structure. In Figure 5, left figure is the inversion result with L2 distance which shows the optimization is trapped at a local minimum and provides errors at the depth between 0.5 to 2.0 km. On the right figure, we perform nonlinear CG algorithm with UOT distance at first 20 iterations and then perform nonlinear CG algorithm with L2 distance at 21-50 iteration. The large-scale information by UOT gradient largely improves the inversion result.

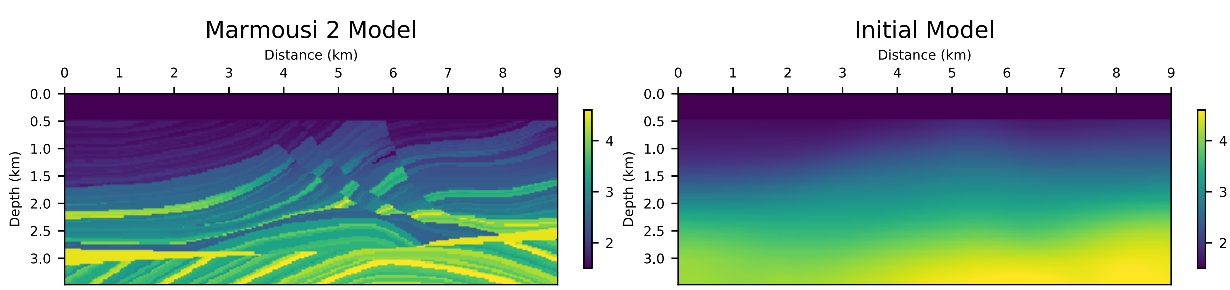

Figure 3. Marmousi 2 model and initial model used in the experiment.

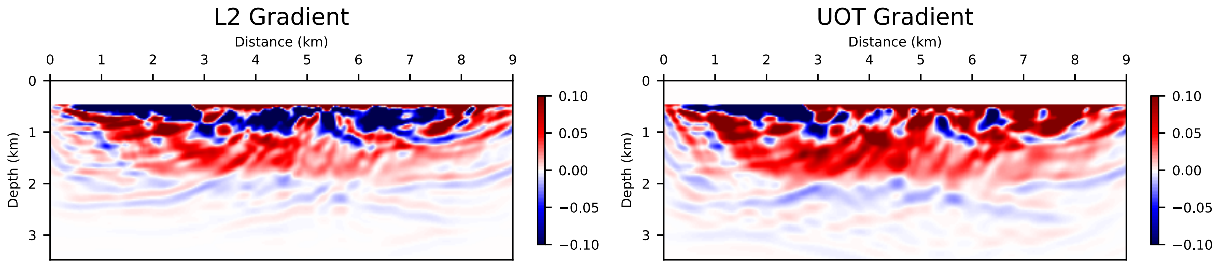

Figure 4. L2 gradient and UOT gradient at the first iteration of optimization.

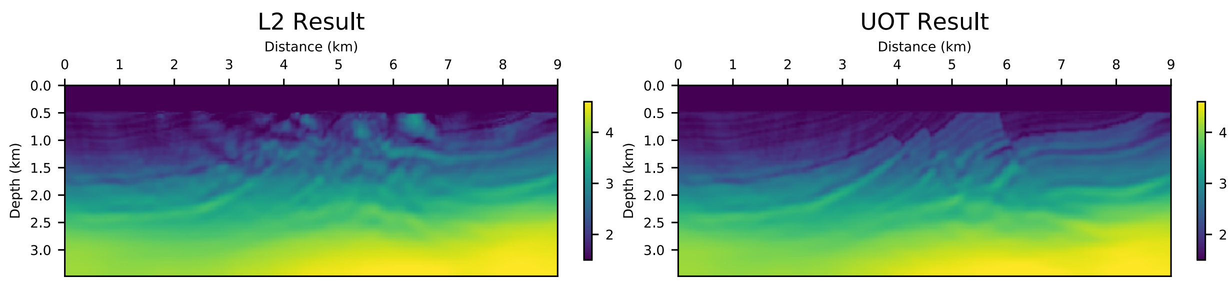

Figure 5. Left figure: L2 inverse result after 50 nonlinear CG iterations. Right figure: inverse result with 1-20 iteration using UOT distance, 21-50 iteration using L2 distance.

Acknowledgements

We gratefully acknowledge the support of CREWES and the industrial sponsors of the CREWES Project. The first author thanks the China Scholarship Council (CSC) for supporting the research. The work of the second author was funded by NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13 and the Discovery grant RGPIN- 2015-06038. The work of the third author was funded by the NSERC Discovery grant RGPIN- 2019-04830.

References

Engquist, B., Froese, B. D., and Yang, Y., 2016, Optimal transport for seismic full waveform inversion: arXiv preprint arXiv:1602.01540.

Métivier, L., Brossier, R., Merigot, Q., Oudet, E., and Virieux, J., 2016a, An optimal transport approach for seismic tomography: Application to 3D full waveform inversion: Inverse Problems, 32, No. 11, 115,008.

Yong, P., Liao, W., Huang, J., Li, Z., and Lin, Y., 2019, Misfit function for full waveform inversion based on the Wasserstein metric with dynamic formulation: Journal of Computational Physics, 399, 108,911.

Yang, Y., Engquist, B., Sun, J., and Hamfeldt, B. F., 2018, Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion: Geophysics, 83, No. 1, R43–R62.

Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X., 2018, Scaling algorithms for unbalanced optimal transport problems: Mathematics of Computation, 87, No. 314, 2563–2609.