Full waveform inversion with unbalanced optimal transport distance

Da Li, Michael P. Lamoureux and Wenyuan Liao

December 11, 2019
1. Optimal transport problem

2. Formulation of full waveform inversion

3. Numerical examples

4. Conclusions
Definition (Optimal transport):
Given $X = Y = \{x_1, x_2, \ldots, x_N\} \subset \mathbb{R}^d$, positive measures $\mu = \sum_i f_i \delta_{x_i}$, $\nu = \sum_i g_i \delta_{x_i}$, with $f_i \geq 0$, $g_i \geq 0$, and $\sum_i f_i = \sum_i g_i$. Let cost matrix $C$ defined by $C_{ij} = |x_i - x_j|^2$. The optimal transport problem between $f$ and $g$ is:

$$\min_{T \in \mathbb{R}^{N \times N}} < T, C > = \sum_{i,j=1}^N T_{ij} C_{ij}, \quad \text{s.t.} \ T 1_N = f, \quad T^T 1_N = g.$$  

The optimal transport distance (2-Wasserstein distance) is given by:

$$W_2^2(f, g) = \min_{T \in \mathbb{R}^{N \times N}} < T, C >, \quad \text{s.t.} \ T 1_N = f, \quad T^T 1_N = g.$$
Optimal transport problem

Limitations:

1. Positive measure: $f_i \geq 0, g_i \geq 0$. For signals, normalization methods are needed.

2. Mass equality condition: $\sum_i f_i = \sum_i g_i$. Unbalanced optimal transport.
Definition (Unbalanced optimal transport):
Given cost matrix $C$, regularization coefficients $\varepsilon$ and $\varepsilon_m$. The unbalanced optimal transport (UOT) distance between $f$ and $g$ is:

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

- Entropy regularization $E(T) = -\sum_{i,j=1}^{N} T_{ij} (\log(T_{ij}) - 1)$, increase the computational efficiency.

- Kullback-Leibler divergence $KL(a|b) = \sum_{i=1}^{N} a_i \left(\log\left(\frac{a_i}{b_i}\right) - 1\right)$ as the mass balancing term.
Theorem (Dual Problem):
Let matrix $K$ defined by $K_{ij} = \exp\left(-\frac{c_{ij}}{\varepsilon}\right)$. The dual problem of equation (1) is given by:

$$W^2_{2,\varepsilon,\varepsilon_m}(f,g) = \max_{\phi,\psi \in \mathbb{R}_N^+} \sum_{i,j=1}^N -\varepsilon_m f_i \left( \exp\left(-\frac{\phi_i}{\varepsilon_m}\right) \right) - \varepsilon_m g_i \left( \exp\left(-\frac{\psi_j}{\varepsilon_m}\right) \right) - \varepsilon K_{ij} \left( \exp\left(\frac{\phi_i}{\varepsilon}\right) \exp\left(\frac{\psi_j}{\varepsilon}\right) - 1 \right).$$

(2)

Strong duality holds. There exists a unique $T^*$ for the equation (1). And $\phi^*, \psi^*$ maximize (2) if and only if

$$T^*_{ij} = \exp\left(\frac{\phi^*_i}{\varepsilon}\right) K_{ij} \exp\left(\frac{\psi^*_j}{\varepsilon}\right).$$
Iterative scaling algorithm:
Starting with an initial value $v^{(0)} = 1_N$, compute iteratively with:

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

until converge requirement is met. Suppose the algorithm converges with $u^*, v^*$, then $\phi^* = \varepsilon \log u^*, \psi^* = \varepsilon \log v^*$. The transport matrix $T^*$ is:

$$T_{ij}^* = u_i^* K_{ij} v_j^*.$$

Moreover, the gradient of the unbalanced optimal transport distance is:

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

For discrete signal $f \in \mathbb{R}^N$, define normalization function $h: \mathbb{R}^N \rightarrow \mathbb{R}^N$.

Two normalization methods are studied:

• Linear normalization: $h_{\text{linear},k}(f) = f + k$.

• Exponential normalization: $h_{\text{exp},k}(f) = \exp(kf)$.

Here $k$ is the normalization coefficient.
Let the spatial domain $\Omega$ large enough. The full waveform inversion is a PDE constrained optimization problem. We use the acoustic wave equation with certain boundary condition as the constraint. Suppose there are $N_s$ sources and $N_r$ receivers in the model.

$$\min_c J[c] = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} W_{z,e,e_m}^2 (h_k(d_{s,r}), h_k(d_{obs,s,r})).$$

s. t. \[ \frac{1}{c^2(x)} \frac{\partial^2}{\partial t^2} u_s(x, t) - \Delta u_s(x, t) = f_s(x, t). \]
\[ d_{s,r}(t) = P_r u_s(x, t) = u_s(x_r, t). \]

The gradient is given by the adjoint state method:

$$\nabla J[c](x) = \sum_{s=1}^{N_s} - \frac{2}{c^3(x)} \left( \frac{\partial^2}{\partial t^2} u_s(x, t) \right) v_s(x, t).$$

Here $v_s$ is the adjoint wavefield.
Adjoint equation:

\[
\frac{1}{c^2(x)} \frac{\partial^2}{\partial t^2} v_s(x, t) - \Delta v_s(x, t) = \tilde{f}_s(x, t).
\]

Where \( \tilde{f}_s \) is the adjoint source:

- **Linear normalization**

\[
\tilde{f}_s = - \sum_{r=1}^{N_r} P_r^T \nabla W^2_{2,\varepsilon,\varepsilon_m} \left( h_k(d_{s,r}), h_k(d_{obs,s,r}) \right).
\]

- **Exponential normalization**: 

\[
\tilde{f}_s = - \sum_{r=1}^{N_r} P_r^T \left( ke^{kd_{s,r}} \right)^T \nabla W^2_{2,\varepsilon,\varepsilon_m} \left( h_k(d_{s,r}), h_k(d_{obs,s,r}) \right).
\]
Two 10 Hz Ricker wavelets $f$ and $g$. The amplitude of $f$ is 1.2 times of $g$.

Define the misfit function:

$$J(s) = d(f(t - s), g(t - 0.5)),$$

Here $d$ is L2 distance, UOT distance with linear and exponential normalization.
Numerical example 1
Velocity model: $c(\delta c, z) = c_0(x, z) + \delta cH(z)$. Here $H(z)$ is the step function. The true model is $c(0.05,0.51)$ as shown in figure (a).

Source position are shown as the red spot. There are 51 receivers on the top of the model.

With homogeneous initial model (b), we compute the misfit function:

$$\hat{J}(\delta c, z) = J[c(\delta c, z)],$$

where $J$ is the misfit function of FWI problem with L2 distance or UOT distance.
Numerical example 2

(a) L2 misfit

(b) UOT (linear) misfit

(c) UOT (exp) misfit
Numerical example 3

- 11 sources with 10 Hz Ricker wavelet on the left-hand side and 101 receivers on the right-hand side.

- Gradient descent method after 5 iterations.
Numerical example 3

(a) L2 adjoint source

(b) UOT (linear) adjoint source

(c) UOT (exp) adjoint source
Conclusions: 
• Comparing to L2 distance, optimal transport distance has better result with respect to time shift. 
• The optimal transport distance provides smooth gradient comparing to L2 distance.

Future works: 
• How to choose the parameters of optimal transport efficiently. 
• Mathematical results on how the optimal transport distance mitigate the cycle-skipping issue. 
• More realistic numerical experiments are needed such as Marmousi 2 model or SEG 2014 benchmark data.
Acknowledgement

- CREWES sponsors
- NSERC
- China scholarship council
Main references

• Yang, Y., and Engquist, B., 2017, Analysis of optimal transport and related misfit functions in full-waveform inversion: Geophysics, 83, No. 1, A7–A12.

• 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.

• Engquist, B., and Froese, B. D., 2013, Application of the Wasserstein metric to seismic signals: arXiv preprint arXiv:1311.4581.

• 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.