A misfit function based on entropy regularized optimal transport for full waveform inversion

Fuqiang Chen *, KAUST and Daniel Peter, KAUST

SUMMARY

Given two probability distributions, the classical theorem of optimal transport aims to determine a transport plan which can map one distribution to the other such that the transport cost is minimized. The general functions such as seismic traces do not satisfy all properties of probability distribution. Therefore, we normalize the seismic traces by an exponential function prior to applying the optimal transport to define the distance between seismic traces. In this abstract, we report some results of full waveform inversion from an alternative misfit function based on entropy regularized optimal transport. The regularization gives a smooth approximation to the original optimal transport and the regularized optimum can be found efficiently though. Numerical examples demonstrate the proposed misfit function can invert the data with super lower frequency unavailable from a rough initial model.

INTRODUCTION

Full waveform inversion is an appealing method to estimate a high resolution model which is supposed to account for the observed data. One of its essential steps is to define a misfit function which is generally believed to contain many local minima and it becomes even worse if the super lower frequency data are missed or we do not know a sufficiently accurate initial model such that cycle skipping issues can be avoided even under the assumption that the exact physics and source mechanism are correctly simulated. These issues limit the application of FWI to field data as the routine technique in velocity model building. There are many efforts have been made to alleviate the restrictions and it remains one of very active research topics.

With the data from surface seismic, the ultimate goal is to characterize the subsurface with model parameters such as velocity and density. FWI iteratively estimates a model by minimizing the data residual. This data fitting problem suffers severely from local minima. With the same data and the same physics, we can also estimate a model in image-domain where the data are accumulated before measuring the difference (Sava, 2014). The latter is generally used to estimate the long wavelength component or the large scale part of a model. One of the common features for velocity estimation in either data-domain or image-domain is that both of them needs to define a misfit function. Therefore, the misfit functions exhibiting less local minima are helpful for both techniques. Many methods other than L_2 norm have been proposed to construct a misfit function such as correlation-based (Van Leeuwen and Mulder, 2010) and envelope-based criterion (Bozdag et al., 2011; Wu et al., 2014). Due to the successful application of optimal transport in other disciplines, the pioneering work applying optimal transport to the seismic signals also demonstrates that misfit function based on optimal transport could reduce the

risk of trapping in a local minima far from global one. Monge formulation in 1D supports the balanced mass transport and the analytical transport plan can be found to define the misfit function for FWI (Engquist and Froese, 2013; Engquist et al., 2016; Yang et al., 2016). Kantorovich formulation is comparatively straightforward to relax the marginal constraints such that the unbalanced mass transport is possible (Métivier et al., 2016). However, it is computationally prohibitive to obtain the exact solution. In this paper, we introduce entropy-regularized optimal transport also named as Sinkhorn distance which admits the smooth approximation to the original problem (Cuturi and Doucet, 2013).

In the following, we begin by reviewing the regularized optimal transport. Then it is physically illustrated with seismic traces. Finally two numerical experiments are conducted to verify the proposed misfit function when the super lower frequency data are not available.

THE OPTIMAL TRANSPORT AS LINEAR PROGRAMMING

In this section, we first review the entropy regularized optimal transport and fill in some missing steps for optimal transport plan calculation. For the proof of more theoretical and properties related to regularized optimal transport, readers can refer to two publications (Cuturi, 2013; Cuturi and Doucet, 2013).

Optimal transport, as its name would suggest, seeks the most economical plan to rearrange one distribution of mass h into the other g such that the cost of transporting all pieces of mass is minimized,

$$W^P(h, g) := \min_{T \in U(h, g)} \langle M, T \rangle, \quad (1)$$

where M denotes the ground metric. $W^P(h, g)$ is referred to as the p-Wasserstein distance if the ground metric is defined by L_p . T denotes a transport plan and $U(h, g)$ is the set of couplings defined as

$$U(h, g) := \left\{ T \in R_+^{n \times n} \mid T \mathbf{1}_n = h, T^\top \mathbf{1}_n = g \right\}, \quad (2)$$

where $g \in R_+^n$, $h \in R_+^n$, $\mathbf{1}_n = (1, 1, \dots, 1, 1)$ is the all-one vector containing n entries. Equation 2 implies that mass is preserved before and after re-distribution if the solution is feasible. Solving equation 1 with constraints 2 by simplex algorithm is cubic complexity. It is still computationally prohibitive even when employing the structure of M . Therefore, many approximate solutions have been proposed. One of them is entropy regularization optimal transport which gives a smooth approximation to the equation 1.

Seismic traces have negative values and the mass of two traces are not equal. Therefore some strategies have been considered, for example, we can separate the signals into positive parts

Full waveform inversion

(h_+, g_+) and negative parts (h_-, g_-) or we can scale the original signals such that they are strictly positive (Engquist and Froese, 2013), then normalize the signals with the same sign individually to ensure that their mass are equal. The former results in the non-smooth signals although it preserves some properties of optimal transport distance. Normalization with scaling does not change the smoothness of signal and which is used in this abstract. Before applying the optimal transport, we apply an exponential scaling to all traces, which reads

$$h(t) \leftarrow e^{ch(t)} / \int_0^{T_0} e^{ch(t)} dt, \quad (3)$$

where c is a constant number to avoid overflow.

Entropy regularization

Solving equation 1 are computational prohibitive even if the solution exists. As an alternative, optimal transport with entropy regularization would give us the smooth approximation (Cuturi, 2013; Cuturi and Doucet, 2013),

$$W_\lambda^P(f, h) := \min_{T \in U(h, g)} \langle M, T \rangle - \frac{1}{\lambda} E(T), \quad (4)$$

where $E(T) = -\sum T_{ij} \log T_{ij}$. Cuturi and Doucet (2013) provided a dual expression for equation 4 as

$$d_\lambda(h, g) := \max_{(\alpha, \beta)} \alpha^\top h + \beta^\top g - \frac{1}{\lambda} \sum_{i,j} e^{-\lambda(m_{ij} - \alpha_i - \beta_j)}, \quad (5)$$

where $\alpha \in \mathbb{R}^n$, $\beta \in \mathbb{R}^n$. Applying first order condition to equation 5, we can have

$$\begin{aligned} h_k &= e^{\lambda \alpha_k} \sum_i e^{-\lambda m_{ki}} e^{\lambda \beta_i} \\ g_k &= e^{\lambda \beta_k} \sum_j e^{-\lambda m_{jk}} e^{\lambda \alpha_j}. \end{aligned} \quad (6)$$

By introducing $u = e^{\lambda \alpha}$, $v = e^{\lambda \beta}$ and $K = e^{-\lambda M}$, we can rewrite equation 6 as,

$$\begin{aligned} h &= u \circ K v \\ g &= v \circ K^\top u, \end{aligned} \quad (7)$$

where $\circ$ is the Hadamard product or entry-wise product. The optimum for the primal equation is given as $T^* = \text{diag}(u) K \text{diag}(v)$ (Cuturi, 2013).

In summary, to apply entropy regularized distance for the construction of misfit function in FWI, the first step is normalizing the seismic traces by equation 3, then calculating the misfit function by either the primal or dual equation. According to the dual equation, we can have α^* as the derivative of misfit function as the first marginal or normalized synthetic data. Practically, we may scale α as well because the fact that $(\alpha + c\mathbf{1}, \beta - c\mathbf{1})$ would be feasible as well if (α, β) is a solution to equation 6. Another concern about misfit function calculation is how we calculate the optimal transport distance between two gathers. Even both the primal and dual equations can be used to calculate the optimal transport plan in higher dimensions, we calculate the distance between two gathers trace by trace (Engquist and Froese, 2013).

Figure 1: Wavelet (a) and its frequency spectrum.

Figure 2: Marmousi model to generate the observed data.

Figure 3: Observed data acquired by solving the acoustic wave equation with the wavelet in Figure 1a. The free surface boundary condition is considered and the full frequency-band random noise are added. The multi-scale inversion is implemented by progressively increasing frequency band: 0~2 Hz (a), 0~3 Hz (b), 0~4 Hz (c) and 0~10 Hz (d). We begin the inversion with 0~3 Hz data.

Full waveform inversion

NUMERICAL EXAMPLES

In this section, we first design a source wavelet shown in Figure 1a and its frequency spectrum is shown in Figure 1b. The true model to be estimated is shown in Figure 2. The observed data is synthesized by solving the acoustic wave equation with free surface boundary condition. 767 receivers are deployed for all 110 shots. For practical applications, exact physics and source mechanism are not exactly known. In addition, observed data contain different types of noise and super lower frequency are not available. In the following, we assume that the physics is correct and source mechanism is known. To show the noise effect on inversion result, we add random noise for all frequency components up to Nyquist and the noise-free data are also inverted as a reference. Figures 3a, 3b, 3c and 3d show the noise-added observed data with different frequency bands. Noting the data under 2 Hz are dominated by random noise and we start the inversion from 3 Hz for both experiments.

To give an illustrative explanation about regularized optimal transport, we extract three traces shown in Figure 4 where the first trace $t_1 = s_1 + n_1$ is the addition of signal s_1 and random noise n_1 and represents the observed data. The second trace t_2 is synthesized with the initial model as shown in Figure 6a. The third trace $t_3 = s_1$ does not contain noise. The last trace $t_4 = n_1$ is the difference between the first and the third to illustrate the noise level. Figures 5b and 5a show the optimal transport plans which can relocate one trace to the other. Figures 6b and 6c are the inversion results from noise-free and noise data.

As we can see that the proposed misfit function can estimate a good model for relatively shallow part even the super-lower frequency data are not reliable because of random noise. For the deep part, especially the part close to the model bottom, the proposed misfit function cannot give an acceptable result even in an ideal world of assuming the physics correct and wavelet known. We expect the reflection-based FWI can improve the deeper part by updating the long-wavelength part with reflection data. On this settings, we still need to define a misfit function for observed reflections and calculated reflections by de-migration process. Therefore, our proposed misfit function could enhance the reflection based FWI as well.

Figure 4: The first three traces are from one observed shot with noise, one synthetic shot and the same observed shot without noise, respectively. The last trace is the difference between the first and the third to indicate the noise level. All traces are filtered by low-pass filter with cutoff frequency 3 Hz.

CONCLUSION

We have proposed a very portable method to enhance the FWI by introducing the entropy regularized optimal transport to define the misfit function. The regularization approximates the original optimal transport but accelerates the solution numerically. To calculate the Fréchet derivative, we start from its dual formulation and show how the misfit and its derivative with respect to the first marginal can be calculated. Numerical examples indicate that the proposed misfit function can invert a model from data with super-lower frequency data unavailable. Misfit function is ubiquitous in seismic inversion such as reflection-based FWI, wave-equation migration velocity analysis and their variants which should be boosted by the proposed misfit function with very little modification for existing codes.

ACKNOWLEDGMENTS

We thank KAUST for providing the powerful computational resources. Most of figures are created by Madagascar, we thank all users and developers who make it friendly to use.

Full waveform inversion

Figure 5: Optimal transport plan between two seismic traces with (a) and without (b) noise.

Figure 6: The initial model which is horizontally invariable and linearly increasing in depth (a), estimated model with noise-free data (b) and with random-noise added data (c).

REFERENCES

- Bozdag, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: *Geophysical Journal International*, **185**, 845–870.
- Cuturi, M., 2013. Sinkhorn distances: Lightspeed computation of optimal transport: in *Advances in Neural Information Processing Systems* 26, 2292–2300.
- Cuturi, M., and A. Doucet, 2013, Fast computation of Wasserstein barycenters: ArXiv e-prints.
- Engquist, B., and B. D. Froese, 2013, Application of the Wasserstein metric to seismic signals: ArXiv e-prints.
- Engquist, B., B. D. Froese, and Y. Yang, 2016, Optimal transport for seismic full waveform inversion: ArXiv e-prints.
- Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016, Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion: *Geophysical Journal International*, **205**, 345–377, <https://doi.org/10.1093/gji/ggw014>.
- Sava, P. C., 2014. A comparative review of wavefield tomography methods.
- Van Leeuwen, T., and W. A. Mulder, 2010, “A correlation-based misfit criterion for wave-equation traveltome tomography: *Geophysical Journal International*, **182**(3), 1383–1394, <https://doi.org/10.1111/j.1365-246X.2010.04681.x>.
- Wu, R.-S., J. Luo, and B. Wu, 2014. Seismic envelope inversion and modulation signal model: *Geophysics*, **79**, WA13–WA24, <https://doi.org/10.1190/geo2013-0294.1>.
- Yang, Y., B. Engquist, J. Sun, and B. D. Froese, 2016. Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion: ArXiv e-prints.