Model recovery below reflectors by optimal-transport FWI

Yunan Yang and Björn Engquist, The University of Texas at Austin

SUMMARY

We analyze different velocity recovery mechanisms of full waveform inversion (FWI) and compare misfit functions based on the least-squares (L^2) norm and optimal transport. The focus will be on recovering velocity even below reflecting interfaces. The inherent nonconvexity of the L^2 norm as misfit leads to many local minima by updating the high-wavenumber components first, which is good for migration but problematic for inversion. On the other hand, the "transport" property behind the optimal transport based misfit functions naturally generates low-wavenumber updates which are ideal for recovering the model kinematics. We will analyze the velocity recovery mechanisms of diving waves and for domains above reflecting layers to see the advantages of using optimal transport type of misfit. Multiple numerical examples with simple layered models as well as a realistic benchmark with salt inclusion demonstrate that optimal-transport FWI is able to update the low-wavenumber components.

INTRODUCTION

Full waveform inversion (FWI) is a computational technique for obtaining high-resolution subsurface properties by minimizing the misfit between observed and synthetic seismic waveforms (Tarantola and Valette, 1982). Unlike the least-squares reverse time migration (LSRTM), FWI also updates the background velocity model, and thus the problem turns into a nonlinear waveform inversion. The mathematical formulation of FWI is a partial differential equation (PDE) constrained optimization. In FWI, the misfit function, i.e., the objective function in the optimization process, is defined as a functional on the data domain. In both time (Tarantola, 1987) and frequency domain (Pratt, 1990; Pratt and Worthington, 1990), the least-squares (L^2) norm has been the most widely used misfit function.

FWI is typically performed using local optimization methods in which the subsurface model is described by using a large number of unknowns, and the number of model parameters is specified a priori (Tarantola, 2005). It is well known that the accuracy of FWI deteriorates from the lack of low frequencies, data noise, and poor starting model. In a sequence of papers, the Wasserstein metrics have been introduced as an alternative to L^2 norm to avoid trapping in FWI optimization due to the cycle skipping problem and to reduce the effect of noise (Engquist and Froese, 2014; Engquist et al., 2016; Métivier et al., 2016; Métivier et al., 2016; Yang and Engquist, 2017, 2018; Yang et al., 2017; Chen et al., 2017). As a useful tool from the theory of optimal transport, the Wasserstein metric computes the minimal cost of rearranging one distribution into another. Methods based on the optimal transport compare the observed and simulated data globally and thus include important phase information.

In our previous studies of the quadratic Wasserstein metric (W_2), we mainly focused on FWI with diving waves or data from shallow reflectors to refine the velocity models. Due to limitations of the acquisition geometry, there might be no diving waves traveling through the depth of interests or being recorded by the receivers. Traditional FWI has difficulty in properly updating deeper features due to the lack of diving waves. Several methods have been proposed recently to better utilize the reflection data in FWI (Ramos-Martinez et al., 2016; Sun et al., 2016; Gomes and Chazalnoel, 2017). In these methods, the high-wavenumber and low-wavenumber components in the gradient of reflection data are separated to update the model kinematics first.

The goal of this paper is to study another important feature of optimal-transport FWI for reflection data. We demonstrate this feature using numerical examples of both simple layered models and a more realistic salt model. Naturally, the inversion improves when diving waves and reflections from shallower interfaces are available. We will see that partial inversion for velocity below the reflectors is still possible by using W_2 norm even when the initial model does not have any waves returning from below the reflector. Finally, based on the observed behavior we conclude that the better convexity of W_2 metric can tackle many different types of local minima that conventional L^2 norm suffers: the cycle-skipping issue and updating only the high-wavenumber components.

THEORY

In this section, we briefly review the mathematical formulas of the adjoint-state method, conventional FWI with L^2 norm and FWI using trace-by-trace W_2 norm.

Without loss of generality, we use a simple acoustic setting in this abstract:

$$\begin{cases} m(\mathbf{x})\frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} - \triangle u(\mathbf{x},t) = s(\mathbf{x},t) \\ u(\mathbf{x},0) = 0 \\ \frac{\partial u}{\partial t}(\mathbf{x},0) = 0 \end{cases} \tag{1}$$

We assume the model $m(\mathbf{x}) = \frac{1}{c(\mathbf{x})^2}$ where $c(\mathbf{x})$ is the velocity, $u(\mathbf{x},t)$ is the wavefield, $s(\mathbf{x},t)$ is the source. It is a linear PDE but a nonlinear operator from model domain $m(\mathbf{x})$ to data domain $u(\mathbf{x},t)$.

Adjoint-State Method

The advances in numerical methods and computational power allow for solving the wave equations and computing the Fréchet derivative with respect to model parameters, which are needed in the optimization. In the adjoint-state method (Plessix, 2006), one only needs to solve two wave equations numerically, the forward propagation and the backward adjoint wavefield prop-

10.1190/segam2018-2998611.1

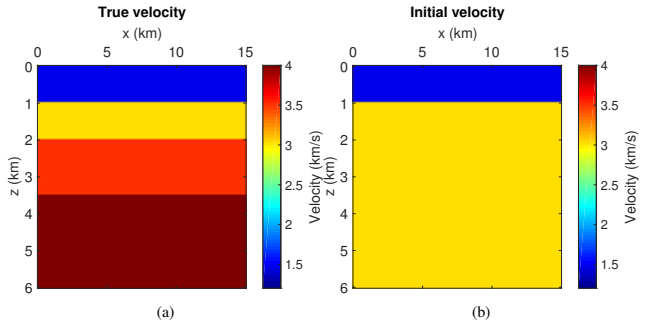


Figure 1: Four-layer model: vertical velocity profiles of (a) the true and (b) the initial model.

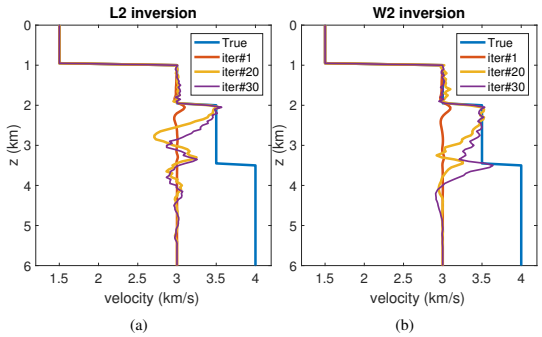


Figure 2: Four-layer model: vertical velocity profiles of (a) L^2 inversion and (b) W_2 inversion after 1, 20 and 30 iterations.

agation:

$$\frac{\partial J}{\partial m} = -\int_0^T \frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} w(\mathbf{x},t)dt, \tag{2}$$

where u is the solution to the forward modelling (1) and w is the solution to the following adjoint wave equation:

$$\begin{cases} m\frac{\partial^2 w(\mathbf{x},t)}{\partial t^2} - \triangle w(\mathbf{x},t) = R^T \frac{\partial J}{\partial f} \\ w(\mathbf{x},T) = 0 \\ w_t(\mathbf{x},T) = 0 \end{cases} \tag{3}$$

FWI with L^2 Norm

In time domain conventional FWI defines a least-squares waveform misfit as

$$J_1(m) = \frac{1}{2}\sum_{r=1}^{R}\int |f(\mathbf{x_r},t;m) - g(\mathbf{x_r},t)|^2 dt, \tag{4}$$

where R is the total number of time history traces, $\mathbf{x_r}$ are receiver locations, g is observed data, and f is simulated data which is part of the source wavefield with model parameter m. The time integral is carried out numerically as a sum. This formulation can also be extended to the case with multiple sources.

FWI with W_2 Metric

Here we review the trace-by-trace technique of using W_2 norm as the misfit function in FWI (Yang et al., 2017):

$$J_2(m) = \sum_{r=1}^{R} W_2^2(f(\mathbf{x_r},t;m), g(\mathbf{x_r},t)), \tag{5}$$

Mathematically it is W_2 metric in the time domain and L^2 norm in the spatial domain.

The functions f and g represent synthetic data and observed data in a corresponding trace. With proper normalization, signals f and g can be rescaled to be positive, supported on $[0,1]$, and have total integral 1. We can exactly solve the 1D optimal transportation problem (Villani, 2003) in terms of the cumulative distribution functions $F(x) = \int_{-\infty}^{x} f(t)\,dt$ $G(x) = \int_{-\infty}^{x} g(t)\,dt$ and their inverses F^{-1} and G^{-1}. The formula for 1D quadratic Wasserstein metric (W_2) between f and g is

$$W_2^2(f,g) = \int_0^1 |F^{-1} - G^{-1}|^2 dy \tag{6}$$

The corresponding Fréchet derivative which is also the adjoint source term in the backward propagation is:

$$\frac{\partial W_2^2(f,g)}{\partial f} = \left(\int_t^{T_0} -2(s - G^{-1}(F(s)) \frac{dG^{-1}(y)}{dy}\Big|_{y=F(s)} f(s)ds\right) dt + |t - G^{-1}(F(t))|^2 dt. \tag{7}$$

This adjoint source term in the discrete 1D setting can be computed as

$$\left[U\ \mathrm{diag}\left(\frac{-2f(t)dt}{g(\varphi(t))}\right)\right](t - \varphi(t))dt + |t - \varphi(t)|^2 dt, \tag{8}$$

where $\varphi = G^{-1} \circ F$ and U is the upper triangular matrix whose non-zero components are 1.

REFLECTIVITY AND SUBLAYER INVERSION

In this section, we will discuss several layered models, whose velocities only vary vertically. The source in all the tests is a Ricker wavelet centered at 15 Hz. There are in total 52 sources and 301 receivers on top in the first (water) layer. Gradients of the water layer are muted in all the tests.

Four-Layer Reflection Model

The Wasserstein metric as the misfit functional in FWI was introduced (Engquist and Froese, 2014) with the motivation to tackle the cycle-skipping issues caused by phase mismatches. However, we will see that W_2 norm is more powerful than phase based inversion since it can invert with reflection only data. One such case is velocity estimation between reflecting discontinuities.

Let us consider a four-layer model (Figure 1a) whose two deepest layers are unknown to the initial model (Figure 1b). The

10.1190/segam2018-2998611.1

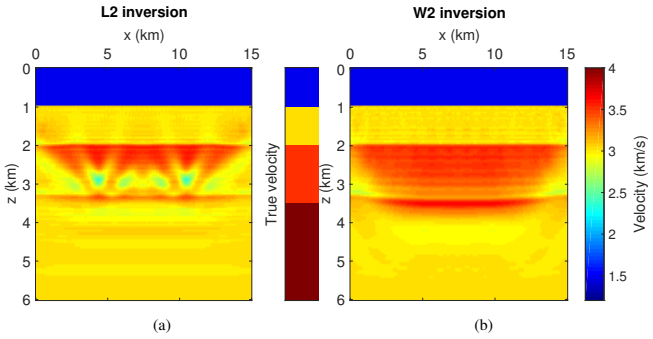


Figure 3: Four-layer model: inversion result of (a) L^2 and (b) W_2 for the four-layer model (Figure 1) after 30 iterations.

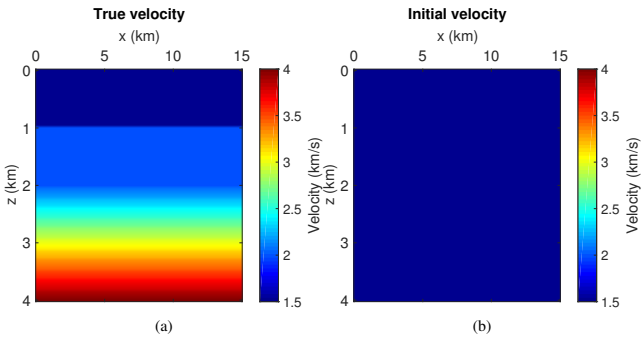


Figure 4: Diving wave guided inversion: vertical velocity profiles of (a) the true model and (b) the initial model.

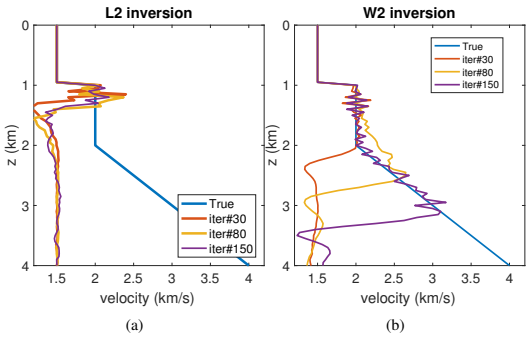


Figure 5: Diving wave guided inversion: 1D velocity profiles of (a) L^2-FWI and (b) W_2-FWI after 30, 80 and 150 iterations.

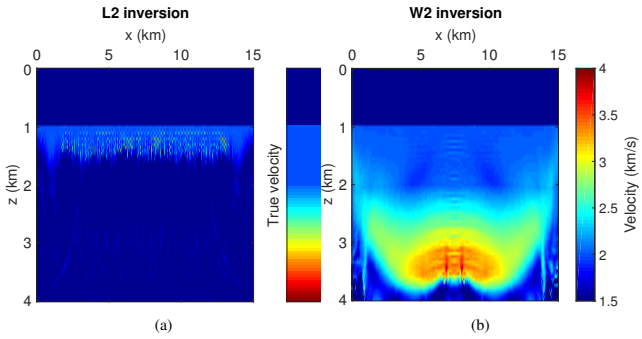


Figure 6: Diving wave guided inversion: FWI using (a) L^2 norm and (b) trace-by-trace W_2 norm after 200 iterations.

total recording time is 4.0 seconds. The observed data only contains a direct wave and multiple reflections from the model discontinuities. The initial data lacks the last reflections from the two deepest interfaces.

In this test, the reflections in the observed data of the unknown layers lead to immediate local velocity corrections by both L^2 and W_2 based FWI, see the first iterations in Figure 2a and Figure 2b. After 30 iterations, L^2-FWI starts to create wrong features as seen in Figure 3a, but W_2-FWI completely recovers the first unknown layer and continues to reconstruct the second one (Figure 3b).

Two facts may contribute to the fast recovery. Not only the velocity in the layer of 3.5 km/s is implicitly embedded in the amplitude of both the last two reflections and the head waves, but also its velocity is interpreted by W_2 norm from the travel time difference between the reflections. In another word, the final reflection generated by the velocity discontinuity jumping from 3.5 km/s to 4 km/s helps W_2 norm to recover the layer of 3.5 km/s significantly.

Diving Wave Guided Inversion

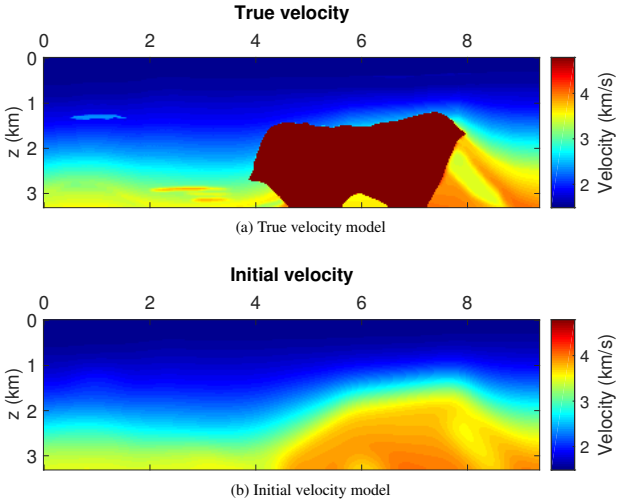
The true model here consists of two layers (1.5 km/s and 2km/s), followed by the linear velocity increasing from 2 km/s to 4 km/s (Figure 4a), while the starting model is a homogeneous velocity 1.5 km/s (Figure 4b). The total recording time is 6.0 seconds. With a gradual velocity increase at the deeper part of the true model, the diving waves are recorded in the observed data besides the reflections.

In the previous four-layer model, two layers are unknown to the initial model, and FWI using W_2 norm quickly recovers the first unknown layer because of the travel time difference between the last two reflections. We can see that in this test travel time between the reflection and the diving wave also help W_2-FWI quickly recover the unknown constant velocity layer in 30 iterations (Figure 5b), and the linearly increasing part is mostly recovered in 200 iterations (Figure 6b). Again, the L^2-FWI is trapped in local minima generated from both the reflections and the diving waves by only updating the high-wavenumber components.

SALT MODEL INVERSION

It is important to investigate the phenomena discussed above in the more realistic settings of well-known velocity models with strongly reflecting salt inclusions. The inversion results from these sharp discontinuities in velocity are quite similar to those

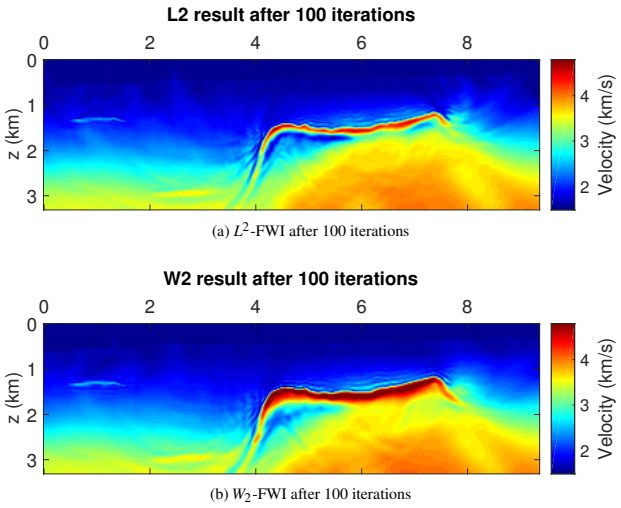
10.1190/segam2018-2998611.1



(a) True velocity model

(b) Initial velocity model

Figure 7: Modified 2004 BP model: (a) the true velocity; (b) the initial velocity.



(a) L^2-FWI after 100 iterations

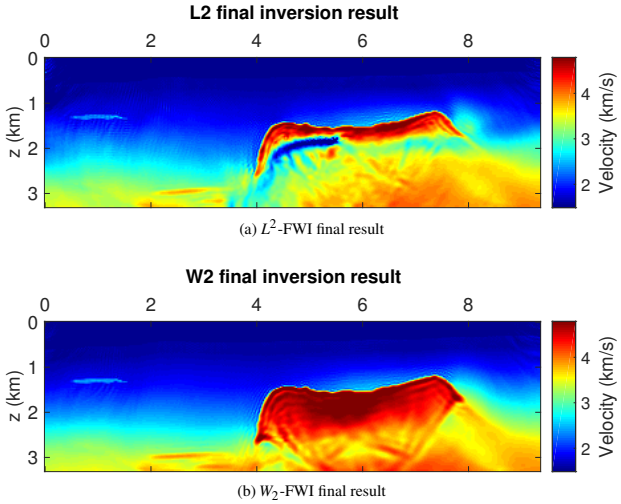
(b) W_2-FWI after 100 iterations

Figure 8: Modified 2004 BP model: (a) FWI with L^2 norm and (b) FWI with trace-by-trace W_2 norm after 100 iterations.

of the simple cases. With many reflections and diving waves contributing to the inversion in the more realistic models, it is, of course, harder to determine the most relevant mechanisms.

Here we consider a salt model, which is part of the 2004 BP benchmark (3.3 km in depth and 9.35 km in width). Figure 7a and Figure 7b show the true and initial velocities. We place 11 sources of 15 Hz Ricker wavelet and 375 receivers equally on the top. The total recording time is 4 seconds. The observed data is dominated by the refection from the top of the salt inclusion.

After 100 iterations, we can see that FWI using both norms can detect the salt upper boundary (Figure 8a and Figure 8b). With more iterations, the W_2 based technique gradually recovers the salt body (Figure 9b), which is much less the case for L^2 norm



(a) L^2-FWI final result

(b) W_2-FWI final result

Figure 9: Modified 2004 BP model: (a) FWI with L^2 norm and (b) FWI with trace-by-trace W_2 norm final results.

as shown in Figure 9a. This also means that features below the salt inclusion can significantly be better determined by using optimal transport related metrics as the mismatch functional, see Figure 4 in (Yang and Engquist, 2017).

CONCLUSION

FWI using least-squares norm is often unable to recover low-wavenumber components of the model because it updates the same type of high-wavenumber components as in data when the initial model is far away from the true model, which may converge to a wrong result. The difficulty in inversion with reflection data is another type of local minima originating from the inherent nonconvexity of the L^2 norm.

One way to tackle this challenge is to consider another misfit function with better convexity. The Wasserstein metric from the optimal transport theory is one such candidate. Optimal transport compares signals globally and naturally considers misfits in both amplitude differences and phase mismatches. Our numerical examples show that the inversion with W_2 norm efficiently avoids the high-wavenumber components update which does not decrease the data misfit in W_2 norm. The better inversion results of W_2 norm demonstrate its capacity to provide better convexity in inversion for many different types of seismic data, including both diving waves and reflections.

ACKNOWLEDGMENTS

We thank Sergey Fomel for helpful discussions and thank the sponsors of the Texas Consortium for Computational Seismology (TCCS) for financial support. This work was also partially supported by NSF DMS-1620396.

10.1190/segam2018-2998611.1

REFERENCES

Chen, J., Y. Chen, H. Wu, and D. Yang, 2018, The quadratic Wasserstein metric for earthquake location: Journal of Computational Physics, **373**, 188–209, https://doi.org/10.1016/j.jcp.2018.06.066.

Engquist, B., and B. D. Froese, 2014, Application of the Wasserstein metric to seismic signals: Communications in Mathematical Sciences, **12**, 979–988, https://doi.org/10.4310/CMS.2014.v12.n5.a7.

Engquist, B., B. D. Froese, and Y. Yang, 2016, Optimal transport for seismic full waveform inversion: Communications in Mathematical Sciences, **14**, 2309–2330, https://doi.org/10.4310/CMS.2016.v14.n8.a9.

Gomes, A., and N. Chazalnoel, 2017, Extending the reach of full-waveform inversion with reflection data. Potential and challenges: 87th Annual International Meeting, SEG, Expanded Abstracts, 1454–1459, https://doi.org/10.1190/segam2017-17731403.1.

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.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016, An optimal transport approach for seismic tomography: Application to 3D full waveform inversion: Inverse Problems, **32**, 115008, https://doi.org/10.1088/0266-5611/32/11/115008.

Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503, https://doi.org/10.1111/j.1365-246X.2006.02978.x.

Pratt, R. G., 1990, Inverse theory applied to multi-source cross-hole tomography. Part II. Elastic wave-equation: Geophysical Prospecting, **38**, 311–329, https://doi.org/10.1111/j.1365-2478.1990.tb01847.x.

Pratt, R. G., and M. Worthington, 1990, Inverse theory applied to multi-source cross-hole tomography. Part I. Acoustic wave-equation method: Geophysical Prospecting, **38**, 287–310, https://doi.org/10.1111/j.1365-2478.1990.tb01846.x.

Ramos-Martinez, J., N. Chemingui, S. Crawley, Z. Zou, A. Valenciano, and E. Klochikhina, 2016, A robust FWI gradient for high-resolution velocity model building: 86th Annual International Meeting, SEG, Expanded Abstracts, 1258–1262.

Sun, D., K. Jiao, X. Cheng, and D. Vigh, 2016, Reflection based waveform inversion: 86th Annual International Meeting, SEG, Expanded Abstracts, 1151–1156.

Tarantola, A., 1987, Inverse problems theory, *in* Methods for data fitting and model parameter estimation: Elsevier.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM.

Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Reviews of Geophysics, **20**, 219–232, https://doi.org/10.1029/RG020i002p00219.

Villani, C., 2003, Topics in optimal transportation, Graduate studies in mathematics 58: American Mathematical Society.

Yang, Y., and B. Engquist, 2017, Analysis of optimal transport related misfit functions in seismic imaging: International Conference on Geometric Science of Information, 109–116.

Yang, Y., and B. Engquist, 2018, Analysis of optimal transport and related misfit functions in full-waveform inversion: Geophysics, **83**, no. 1, A7–A12, https://doi.org/10.1190/geo2017-0264.1.

Yang, Y., B. Engquist, J. Sun, and B. D. Froese, 2017, Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion: Geophysics, **83**, no. 1, 1–103. https://doi.org/10.1190/geo2016-0663.1.