DOI:10.1190/segam2019-3216471.1

Improving optimal transport based FWI through data normalization

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

SUMMARY

Optimal transport based misfit functions were developed to improve full-waveform inversion (FWI) by reducing the so-called cycle-skipping issues. There have been successful applications on synthetic examples and field data inversions, but there are not enough explanations for why it works. We will present some analysis by analyzing the preparation or normalization of seismic data for the optimal transport step and discuss its connections to many useful features of other techniques in seismic inversion as, for example, the Huber norm, low-frequency enhancement, increased regularity and model extension. The analysis is also used for further improving optimal transport based FWI.

INTRODUCTION

At the heart of seismic exploration is the estimation of essential geophysical properties including wave velocity. The computational technique referred to as full-waveform inversion (FWI) (Tarantola and Valette, 1982; Lailly, 1983) utilizes information of the entire wavefield and follows the standard strategy of a partial differential equation (PDE) constrained optimization. Currently, FWI can reconstruct sub-surface parameters with stunning details and resolution (Virieux et al., 2014).

In both time (Tarantola and Valette, 1982) and frequency (Pratt and Worthington, 1990) domains, the least-squares norm (ℓ^2) has been the most widely used misfit function. It is, however now well known that inversion techniques based on ℓ^2 face three critical obstacles. First, the accuracy of ℓ^2-based FWI is severely hampered by the lack of low-frequency data and a poor starting model. Second, in addition to the difficulties with local minima, an additional problem of the ℓ^2 norm is exacerbated by the fact that observed signals usually suffer from noise in the measurements. Third, traditional FWI has difficulty in accurately updating deeper features with reflection-dominated data.

Current challenges of nonlinear inverse problems motivate us to replace the conventional ℓ^2 norm with a new metric of better convexity and stability. Engquist and Froese (2014) first proposed to use the Wasserstein distance as an alternative misfit function measuring the differences between synthetic data f and observed data g to mitigate the sensitivity to noise and the non-convexity issues of FWI. Soon after the proposal (Engquist and Froese, 2014), there have been fruitful activities from both academia (Engquist et al., 2016; Métivier et al., 2016; Métivier et al., 2016; Yang et al., 2018; Chen et al., 2017; Ballesio et al., 2018; Motamed and Appelö, 2018) and industry (Qiu et al., 2017; Ramos-Martínez et al., 2018; Poncet et al., 2018) in developing the idea of using optimal transport based metrics for inverse problems. Many exciting and encouraging inversions with real data from the oil field have also come out within the past three years. One of the optimal transport based objective functions is to compare seismic signals trace by trace using the quadratic Wasserstein distance (W_2):

$$J_1(m) = \sum_{r=1}^{R} W_2^2(f(\mathbf{x_r},t;m), g(\mathbf{x_r},t)), \tag{1}$$

where $f(\mathbf{x_r},t)$ and $g(\mathbf{x_r},t)$ denote the normalized synthetic data and observed data at each fixed spatial location $\mathbf{x_r}$ and R is the total number of receivers. One can also compare the entire datasets without fixing the spatial location by solving higher-dimensional optimal transport problems.

In optimal transport theory, there are two main requirements for signals f and g compactly supported on domain Ω:

$$f \geq 0,\ g \geq 0,\ < f >= \int_\Omega f = \int_\Omega g =< g > . \tag{2}$$

Since these constraints are not expected for seismic signals, some data pre-processing is needed before we can implement the Wasserstein-based FWI. In Yang et al. (2018), we normalized the signals by scaling and adding a constant:

$$\tilde{f} = \frac{f+c}{< f+c >},\ \tilde{g} = \frac{g+c}{< g+c >},\ c > 0. \tag{3}$$

An exponential based normalization was proposed in Qiu et al. (2017), which we here generalize by adding a constant:

$$\tilde{f} = \frac{\exp(bf)+c}{< \exp(bf)+c >},\ \tilde{g} = \frac{\exp(bg)+c}{< \exp(bf)+c >},\ b,c \geq 0. \tag{4}$$

Although the normalization (3) does not give a convex misfit functional with respect to simple shifts (Yang and Engquist, 2018), it works remarkably well in realistic large-scale examples (Yang et al., 2018). Earlier the constant c in (3) was just to guarantee a positive function and in (4) c was set to zero. These empirical observations motivate us to study further the positive influences of the data normalization on optimal transport based FWI, which we treat as PDE-constrained optimization. We will here focus on (4) for which convexity with respect to shifts can be achieved for sets of b and c values. With experience from realistic numerical tests and the analysis in this paper, it is clear that adding positive c has extended benefits.

In this abstract, we discuss and analyze several advantages of normalization (4) for optimal transport based objective functions in FWI. The normalization allows us to apply the Wasserstein distance to signed signals which can be seen as a type of unbalanced optimal transport. The most important feature is that these normalization methods turn the optimal transport objective function into a "Huber-type" norm. Researchers have studied the robustness of the Huber norm (Huber et al., 1973) in FWI over the years (Guitton and Symes, 2003; Ha et al., 2009; Brossier et al., 2010) which combines the best properties of ℓ^2 squared loss and ℓ^1 absolute loss by being strongly

10.1190/segam2019-3216471.1

convex when close to the target/minimum and less steep for extreme values or outliers. Besides, we will show that (4) helps smooth the optimal transport map between the synthetic and objective data and consequently enhance the low-frequency contents in the adjoint source for back-propagation as well as extends the functional space the optimal map lies.

BACKGROUND

In this section, we briefly review the mathematical formulations of FWI using the quadratic Wasserstein distance (W_2) as the objective function. Without loss of generality, we will introduce all these methods in a simple acoustic setting, but there is ongoing work of using Wasserstein-based objective functions for VTI model as well as isotropic elastic inversions.

Full-Waveform Inversion

We denote the modeled synthetic data as $f(\mathbf{x_r},t;m) = u(\mathbf{x_r},z=0,t;m)$, where $\mathbf{x_r}$ is the receiver location, u is the wavefield, obtained by solving, for example, the following 2D acoustic wave equation:

$$\begin{cases} m(\mathbf{x})\frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} - \triangle u(\mathbf{x},t) = s(\mathbf{x},t), \\ u(\mathbf{x},0) = 0, \quad \frac{\partial u}{\partial t}(\mathbf{x},0) = 0, \end{cases} \tag{5}$$

where the model parameter is the squared slowness $m(\mathbf{x}) = \frac{1}{c(\mathbf{x})^2}$ where $c(\mathbf{x})$ is the wave velocity, $u(\mathbf{x},t)$ is the wavefield and $s(\mathbf{x},t)$ is the source. The above wave equation can be seen as the forward operator F such that $f = F(m)$. The goal of FWI is to recover the subsurface model parameters m^* from the observed true data g such that $F(m^*) = g(\mathbf{x_r},t)$. It estimates the true model parameter m^* through the solution of an optimization problem

$$m^* = \underset{m}{\operatorname{argmin}} J(F(m),g), \tag{6}$$

where J is a suitable choice of objective function.

Due to the efficiency of the adjoint-state method (Plessix, 2006), one only needs to solve two wave equations numerically, the forward propagation and the backward adjoint wavefield propagation, to obtain the gradient for all variables. Different misfit functions typically only affect the source term in the adjoint wave equation. The gradient is

$$\frac{\partial J}{\partial m} = -\int_0^{T_0} \frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} w(\mathbf{x},t)dt, \tag{7}$$

where u is the solution to the forward modelling (5) and w is the solution to the adjoint wave equation with source $\frac{\partial J}{\partial f}$, which is the Fréchet derivative of the objective function with respect to the synthetic data.

The Wasserstein Distance

The optimal mass transport problem seeks the most efficient way of transforming one distribution of mass to the other, relative to a given cost function. It was first brought up by Monge (1781). The optimal transport cost, which is also called the Wasserstein distance, is a class of well-defined metrics (Villani, 2003):

Definition 1 (The Wasserstein distance)**.** *We denote by $\mathscr{P}_p(X)$ the set of probability measures with finite moments of order p. For all $p \in [1,\infty)$ and $\mu,\nu \in \mathscr{P}_p(X)$*

$$W_p(\mu,\nu) = \left(\inf_{T_{\mu,\nu}\in\mathscr{M}} \int_{\mathbb{R}^d} \left|x - T_{\mu,\nu}(x)\right|^p d\mu(x)\right)^{\frac{1}{p}}. \tag{8}$$

$\mathscr{M}$ is the set of all maps that rearrange μ into ν.

For FWI, one can use the quadratic Wasserstein metric ($p = 2$) to measure the misfit in the time domain, and the ℓ^2 norm for the spatial domain as objective function (1). Computing the misfit function becomes a 1D optimal transport problem, which can be solved explicitly:

$$W_2^2(f,g) = \int_0^{T_0} |t - G^{-1}(F(t)|^2 f(t)dt, \tag{9}$$

The corresponding Fréchet derivative (Yang et al., 2018) which is also the adjoint source term in the backward propagation is

$$\frac{\partial W_2^2(f,g)}{\partial f} = \left(\int_t^{T_0} -2(s - G^{-1}(F(s))\frac{dG^{-1}(y)}{dy}\Big|_{y=F(s)} f(s)ds\right) + |t - G^{-1}(F(t))|^2. \tag{10}$$

where F and G are cumulative distribution functions of normalized signals f and g. Chain rule is applied to compute the Fréchet derivative with respect to the raw data.

PROPERTIES

Consider f and g are time histories of raw seismic signals and $\tilde{f}$ and $\tilde{g}$ are normalized probability densities under scaling method (4). We use $W_2(\tilde{f},\tilde{g})$ as the objective function measuring the misfit between raw seismic signals f and g. It can also be viewed as a new loss function W with hyperparameters b and c as follows:

$$W(f,g;b,c) = W_2(\tilde{f},\tilde{g}). \tag{11}$$

Huber-Type Regularization

The Huber norm has better performance in reconstructing different parts of the model in FWI than the traditional least-squares method, in particular, for noise-contaminated data because the Huber norm penalizes much less for large errors than the ℓ^2 norm. For example, when comparing two shifted signals and the error is set as the shift, the Huber norm is $\mathscr{O}(s^2)$ for small shift s while becomes $\mathscr{O}(s)$ for s larger than a threshold. Different ways have been proposed in the literature to define the threshold for the transition (Huber et al., 1973; Guitton and Symes, 2003; Ha et al., 2009).

Next, we will demonstrate that the linear normalization (3) together with the quadratic Wasserstein distance yields a "Huber-type" norm when measuring 1D probability distributions. The threshold for the transition between ℓ^1 and ℓ^2 depends on the constant c.

10.1190/segam2019-3216471.1

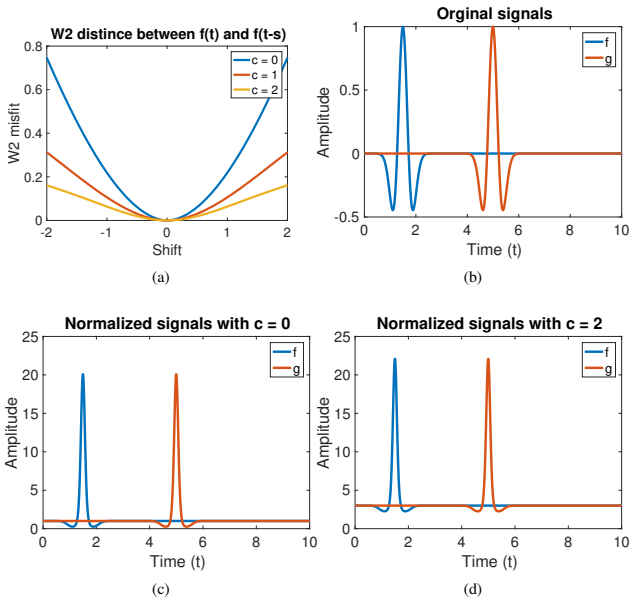

Figure 1: (a) The Huber effect of linear constant c on the loss function $W(f,g;c)$ as a function of shift s; (b) the raw signal f and $g = f(t-s)$; (c) normalized signals $\tilde{f}$ and $\tilde{g}$ with $c = 0$ in (4); (d) normalized signals $\tilde{f}$ and $\tilde{g}$ with $c = 2$ in (4).

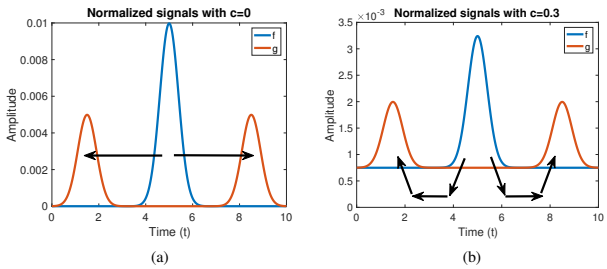

Figure 2: The black arrows represent the optimal map T. (a) Signals f and g are compactly supported on the domain and the optimal map is discontinuous at $t = 5$. (b) Signals f and g after linear normalization. The optimal map T becomes a continuous function by adding the constant $c = 1$.

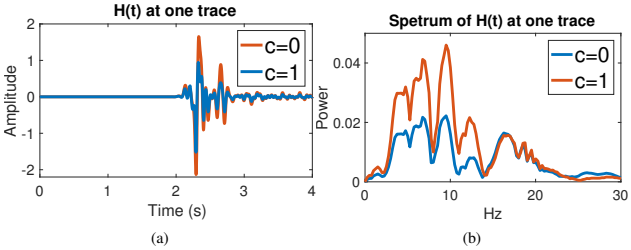

Figure 3: (a) One trace of the adjoint sources of W_2-FWI for the Marmousi model. The blue plot is for the case of $c = 0$ and the red plot is for linear normalization with $c = 1$. (b) The spectrum of the adjoint sources at one trace with different normalization schemes.

Theorem 1 (Huber-Type Regularization)**.** *Let f and g be probability density functions compactly supported on $\Omega \subseteq \mathbb{R}$ and $g(x) = f(x-s)$. Consider $\tilde{f}$ and $\tilde{g}$ as new density functions defined by linear normalization* (3) *for a given $c > 0$, then*

$$W_2^2(\tilde{f},\tilde{g}) = \begin{cases} \mathcal{O}(|s|^2), & \text{if } |s| \le \frac{1}{c} + |\operatorname{supp}(f)|. \\ \mathcal{O}(|s|), & \text{otherwise.} \end{cases} \tag{12}$$

Proof. Without loss of generality, we assume f is compactly supported on $[a_1,a_2]$ and $s \ge 0$. Based on the 1D explicit formula (1), one can compute

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

$$\begin{cases} 2\int_{y_1}^{y_3} |F^{-1}(y) - \frac{y}{c_1}|^2 dy + \int_{y_3}^{y_2} |F^{-1}(y) - F^{-1}(y - c_1 s) + s|^2 dy, \\ \text{if } |s| \le \frac{1}{c} + |a_2 - a_1|, \\ \\ 2\int_{y_1}^{y_2} |\tilde{F}^{-1}(y) - \frac{y}{c_1}|^2 dy + \int_{y_2}^{y_3} \frac{1}{c^2} dy, \text{ otherwise.} \end{cases}$$

Here $F, G, \tilde{F}, \tilde{G}$ are cumulative distribution functions of $f, g, \tilde{f}, \tilde{g}$ respectively, $c_1 = \frac{c}{1+c|\Omega|}$, $y_1 = c_1 a_1$, $y_2 = c_1 a_2 + \frac{1}{1+c|\Omega|}$ and $y_3 = c_1 a_1 + c_1 s$. Since y_1 and y_2 are independent of s, one can show by calculus that $W_2^2(\tilde{f},\tilde{g})$ is linear in s if $s > \frac{1}{c} + |a_2 - a_1|$ and $W_2^2(\tilde{f},\tilde{g}) = \mathcal{O}(s^2)$ if $0 \le s \le \frac{1}{c} + |a_2 - a_1|$. □

Unlike the 1-Wasserstein distance (W_1) which corresponds to the case of $p = 1$ in (8), $W_2^2(f,g)$ is not invariant under mass addition/subtraction in a sense that $W_1(f+c, g+c) = W_1(f,g)$ and $W_2^2(f+c, g+c) \le W_2^2(f,g)$ for given $c \ge 0$. Based on the subadditivity of W_2 under rescaled convolution (Villani, 2003),

$$W_2^2(\tilde{f},\tilde{g}) = W_2^2\left(\frac{f+c}{1+c|\Omega|}, \frac{g+c}{1+c|\Omega|}\right) \le \frac{W_2^2(f,g)}{(1+c|\Omega|)^2}, \tag{13}$$

where $|\Omega|$ denotes the Lebesgue measure of the bounded domain Ω. The inequality shows that the linear normalization decreases loss computed by the original objective function and is the reason for the Huber-type behavior.

To better illustrate the "Huber-type" regularization, we consider signals f and g where f is a single Ricker wavelet and $g = f(t-s)$, see Figure 1b. Since seismic signals are not probability densities, for better convexity data normalization (4) is applied to attain $\tilde{f}$ and $\tilde{g}$ instead of (3). Under proper choice of the hyperparameter b, the Huber-type regularization proved in Theorem 1 still applies. With b fixed, Figure 1a shows the optimization landscape of the objective function with respect to shift s for different choices of c. By comparing plots in Figure 1a, we observe that as c increases, the objective function $W(f,g;b,c)$ becomes less quadratic and more linear, which is also demonstrated in Theorem 1. Figure 1c shows the normalized Ricker wavelets for $c = 0$ while Figure 1d shows the data with constant $c = 2$ applied in (4).

10.1190/segam2019-3216471.1

The Gradient-Smoothing Property

In addition to the fact that normalization (4) turns W_2 based objective function into a "Huber-type" loss, it changes the structure, and more importantly, the regularity of the optimal map T. Based on the adjoint-state method briefly explained in (7), different choice of objective function mainly affects the model gradient through the adjoint source term $\frac{\partial J}{\partial f}$ in the adjoint-state equation.

In the 1D technique (1), the adjoint source has an explicit formula (10) which highly depends on the optimal map $T = G^{-1} \circ F$. Simple calculation demonstrates that the smoothness of the optimal map directly determines the frequency spectrum of the W_2 adjoint source $H(t) = \frac{\partial W_2^2(f,g)}{\partial f}$ as

$$\frac{dH}{dt} = 2\,(t - G^{-1}F(t)) = 2(t - T(t)). \tag{14}$$

Due to non-local features of optimal transport problem, the optimal map is often discontinuous even if $f, g \in C^\infty$; see Figure 2a for example in which the corresponding cumulative distribution functions F and G are monotone but not strictly monotone. However, by adding a nonzero constant to the signals, we make F and G to be strictly monotone and thus invertible in the classical sense. The optimal map also becomes more smooth as Figure 2b shows. Consider $f, g \in C^p$, then $F, G, F^{-1}, G^{-1} \in C^{p+1}$ and hence $T \in C^{2p+1}$. Note that it is not true if G is not strictly monotone, but adding a constant could alter that. Figure 3 shows one example that T itself has higher regularity and the adjoint source has more low frequencies after setting $c = 1$ in the data normalization (4).

Adding a constant in the data normalization also embeds a flavor of the "Extended Modeling" (Symes, 2008). Given x_1, x_2 and $x_2 = T_1(x_1)$, the transport which was originally done by one map T_1 now extends to the composition of a series of maps: $x_2 = T_2 \circ T_2 \circ \cdots \circ T_2(x_1)$; see Figure 2 for illustration. Here the model extension is done within the calculation of the Wasserstein distance. One of the advantages is to enlarge the functional space where the map T lies, for example, from L^∞ in Figure 2a to C^∞ in Figure 2b. A larger functional space increases the chance of a nonempty projection of the adjoint source onto the wave equation constraint, which guarantees a feasible gradient in each iteration and a continuous convergent path from the initial model to the target model in the iterative inversion process.

NUMERICAL EXAMPLE

In this section, we invert the Marmousi model with (1) as the objective function. Figure 4a shows the true model and the inversion starts with initial model in Figure 4b which only varies vertically. We place 21 evenly spaced sources on top at 50 m depth and 307 receivers on top at the same depth with a 30 m fixed acquisition. The total recording time is 4 seconds. The source is a 5 Hz Ricker wavelet. We normalize the data as (4) with fixed b and different choices of c. The inversion stops after a few iterations with no feasible descent direction in the case of $c = 0$ as Figure 4c shows while adding the constant $c = 1$ yields a much better inversion and Figure 4d shows the final inversion result. It can be further improved by continuing with high-frequency sources. The normalization method (4) have several advantages including the Huber-type regularization, the gradient-smoothing property and consequently has a direct improvement for optimal transport based inversion.

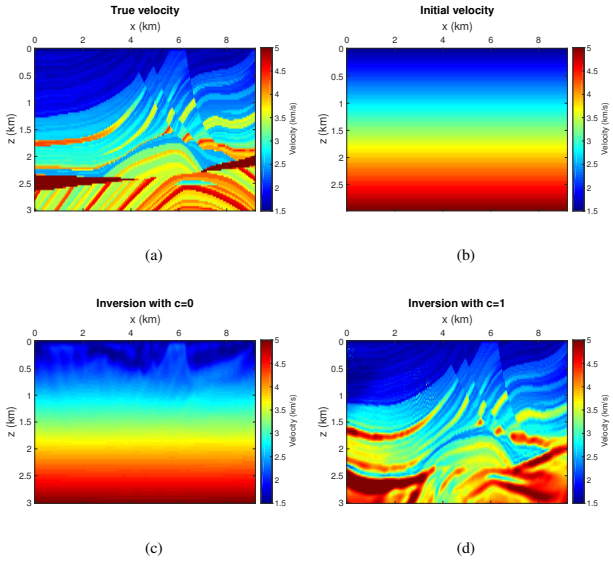

Figure 4: (a) True velocity and (b) inital velocity for the Marmousi model; (c) W_2 based inversion with $c = 0$ in normalization (4) and (d) with $c = 1$ in normalization (4)

CONCLUSION

In this paper, we study the effects of adding a positive constant in the data normalization as preconditioning of optimal transport based FWI. We show that this simple modification introduces several useful features that are well known for other FWI techniques. The resulting misfit function can be seen as a Huber-type norm with high regularity and enhanced lower-frequency content. There is also a clear connection to extended modeling. Most of these properties are also valid for linear scaling (3). The analysis here can explain the earlier contradictory observations (Yang et al., 2018) that linear normalization often worked better in applications than other scaling methods even if it lacks convexity with respect to shifts Engquist and Yang (2018). Our focus here has been on exponential scaling with added constant for which the misfit is also convex as a function of shifts.

ACKNOWLEDGMENTS

We thank Dr. Lingyun Qiu 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.

REFERENCES

Ballesio, M., J. Beck, A. Pandey, L. Parisi, E. von Schwerin, and R. Tempone, 2018, Multilevel Monte Carlo acceleration of seismic wave propagation under uncertainty: arXiv preprint arXiv:1810.01710.

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

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

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

Engquist, B., and Y. Yang, 2018, Seismic inversion and the data normalization for optimal transport: arXiv preprint arXiv:1810.08686.

Guitton, A., and W. W. Symes, 2003, Robust inversion of seismic data using the Huber norm: Geophysics, **68**, 1310–1319, doi: https://doi.org/10.1190/1.1598124.

Ha, T., W. Chung, and C. Shin, 2009, Waveform inversion using a back-propagation algorithm and a Huber function norm: Geophysics, **74**, no. 3, R15–R24, doi: https://doi.org/10.1190/1.3112572.

Huber, P. J., 1973, Robust regression: Asymptotics, conjectures and Monte Carlo: The Annals of Statistics, **1**, 799–821, doi: https://doi.org/10.1214/aos/1176342503.

Kantorovich, L. V., 2006, On the translocation of masses: Journal of Mathematical Sciences, **133**, 1381–1382, doi: https://doi.org/10.1007/s10958-006-0049-2.

Lailly, P., and J. B. Bednar, 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering: Theory and Application, 206–220.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016a, Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion*:* Geophysical Journal International, **205**, 345–377, doi: https://doi.org/10.1093/gji/ggw014.

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

Monge, G., 1781, Mémoire sur la théorie des déblais et des remblais: Histoire de l'Académie Royale des Sciences de Paris.

Motamed, M., and D. Appelo, 2018, Wasserstein metric-driven Bayesian inversion with application to wave propagation problems: arXiv preprint arXiv:1807.09682.

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, doi: https://doi.org/10.1111/j.1365-246x.2006.02978.x.

Poncet, R., J. Messud, M. Bader, G. Lambaré, G. Viguier, and C. Hidalgo, 2018, FWI with optimal transport: A 3D implementation and an application on a field dataset: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: https://doi.org/10.3997/2214-4609.201801029.

Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multi-source cross-hole tomography — Part 1: Acoustic wave-equation method: Geophysical Prospecting, **38**, 287–310, doi: https://doi.org/10.1111/j.1365-2478.1990.tb01846.x.

Qiu, L., J. Ramos-Martínez, A. Valenciano, Y. Yang, and B. Engquist, 2017, Full-waveform inversion with an exponentially encoded optimal-transport norm: 87th Annual International Meeting, SEG, Expanded Abstracts, 1286–1290, doi: https://doi.org/10.1190/segam2017-17681930.1.

Ramos-Martínez, J., L. Qiu, J. Kirkebø, A.A. Valenciano, and Y. Yang, 2018, Long-wavelength FWI updates beyond cycle skipping: 88th Annual International Meeting, SEG, Expanded Abstracts, 1168–1172, doi: https://doi.org/10.1190/segam2018-2998433.1.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790, doi: https://doi.org/10.1111/j.1365-2478.2008.00698.x.

Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Reviews of Geophysics, **20**, 219–232, doi: https://doi.org/10.1029/rg020i002p00219.

Villani, C., 2003, Topics in optimal transportation: Graduate Studies in Mathematics 58.

Virieux, J., A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, 2017, An introduction to full waveform inversion, *in* Encyclopedia of exploration geophysics: SEG, R1–1.

Yang, Y., and B. Engquist, 2017, Analysis of optimal transport and related misfit functions in full-waveform inversion: Geophysics, **83**, no. 1, A7–A12, doi: https://doi.org/10.1190/geo2017-0264.1.

Yang, Y., B. Engquist, J. Sun, and B. F. Hamfeldt, 2018, Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion: Geophysics, **83**, no. 1, R43–R62, doi: https://doi.org/10.1190/geo2016-0663.1.