Analysis of optimal transport and related misfit functions in FWI

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

SUMMARY

We summarize and compare four different misfit functions for full waveform inversion (FWI): the conventional least-squares norm, the integral wavefields misfit functional, the Normalized Integration Method (NIM) and the quadratic Wasserstein metric. The integral wavefields misfit functional and NIM are equivalent to the norm for Soblev space, which has intrinsic connections with the quadratic Wasserstein metric. We extract two important features of optimal transport. The first one is integration of data, which reduces high frequencies and globally compares observed and synthetic seismic waveforms. The other is rescaling of the data to be nonnegative. Numerical results illustrate that FWI with quadratic Wasserstein metric can effectively overcome the cycle skipping problem. A mathematical study on the convexity of the four misfit functions demonstrates the importance of data nonnegativity and integration in dealing with local minima in inversion.

INTRODUCTION

Full waveform inversion (FWI) is a data-driven method to obtain high resolution subsurface properties by minimizing the difference between observed and synthetic seismic waveforms (Virieux et al., 2017). In the past three decades, the least-squares norm (L^2) has been widely used as a misfit function (Tarantola and Valette, 1982; Lailly, 1983), which is known to suffer from cycle skipping issues with local minimum trapping and sensitivity to noise (Virieux and Operto, 2009). Other misfit functions proposed in literature, include the L^1 norm (Brossier et al., 2010), the Huber norm (Ha et al., 2009), filter based misifit functions (Warner and Guasch, 2014; Zhu and Fomel, 2016), seismic envelop inversion (Luo and Wu, 2015) and some others.

A recently introduced class of misfit functions are optimal-transport related (Engquist and Froese, 2014; Métivier et al., 2016; Engquist et al., 2016; Métivier et al., 2016; Yang et al., 2016). As useful tools from the theory of optimal transport, the quadratic Wasserstein metric (W_2) computes the optimal cost of rearranging one distribution into another with a quadratic cost function, while 1-Wasserstein metric (W_1) using absolute value cost function.

In this paper, we will also discuss about Normalized Intergration Method (NIM) which computes the least-squares difference between two normalized data sets (Liu et al., 2012; Chauris et al., 2012; Donno et al., 2013). If we consider the data are properly rescaled, the misfit of NIM is the norm of Sobolev space H^{-1} in mathematics. The connection between W_2 and H^{-1} is not obvious from the optimal transport definition, but is clear from the 1D closed solution formula. We shall also see that this is valid in higher dimensions even if there is no explicit solution formula.

The goal of this paper is to analyze important features of optimal transport and to compare with methods introduced earlier. We focus on two features in particular. One is integration of data and the other is the need to rescale the data to be nonnegative. Integration provides a global comparison between observed and synthetic data and also shifts the focus to lower frequencies. Nonnegativity further reduces the risk of cycle skipping.

THEORY

Full waveform inversion is a PDE-constrained optimization problem, minimizing the data misfit $d(f,g)$ by updating the model m, i.e. :

$$m^\star = \underset{m}{\operatorname{argmin}}\ d(f(x_r,t;m),g(x_r,t)), \tag{1}$$

where g is observed data, f is simulated data, x_r are receiver locations, and m is the model parameter. We get the modeled data $f(x,t;m)$ by solving a wave equation with a finite difference method (FDM) in both the space and time domain (Alford et al., 1974).

Generalized least squares functional is a weighted sum of the squared errors and hence a generalized version of the standard least squares misfit function. The formulation is

$$J_1(m) = \sum_r \int |W(f(x_r,t;m)) - W(g(x_r,t))|^2\, dt, \tag{2}$$

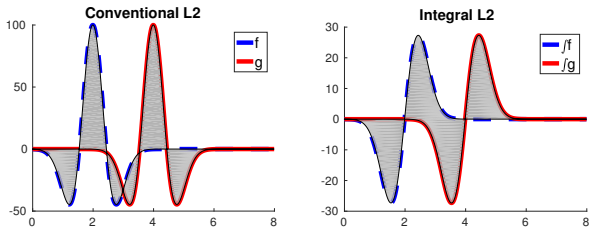
where W is an operator. In the conventioinal L^2 misfit, $W = I$, the identity operator.

The integral wavefields misfit functional (Huang et al., 2014) is a generalized least squares functional applied on full-waveform inversion (FWI) with weighting operator $W(u) = \int_0^t u(x,\tau)d\tau$. The objective function is defined as

$$J_2(m) = \sum_r \int \left| \int_0^t f(x_r,\tau;m)d\tau - \int_0^t g(x_r,\tau)d\tau \right|^2 dt, \tag{3}$$

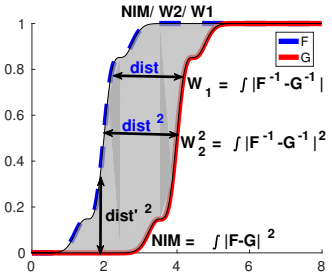
If we define the integral wavefields $U(x,t) = \int_0^t u(x,\tau)d\tau$, then misfit function (3) is the ordinary least squares misfit between the observed and predicted integral wavefields $\int_0^t g(x_r,\tau)d\tau$ and $\int_0^t f(x_r,\tau;m)d\tau$. The integral wavefields still satisfy the original acoustic wave equation with a different source term: $\delta(\vec{x}-\vec{x}_s)\int_0^t s(\tau)d\tau = \delta(\vec{x}-\vec{x}_s)H(t)*s(t)$, where s is the original source term and $H(t)$ is the Heaviside step function (Huang et al., 2014).

Normalized Integration Method (NIM) is another generalized least squares functional, similar to the integral wavefields misfit functional. However, compared with integral wavefields misfit functional which directly integrates the observed and



(a) Misfit for L^2 norm

(b) Misfit for Integral L^2 method



(c) Comparison among NIM, W_2, W_1 in 1D

Figure 1: The shaded areas represent the mismatch each misfit function considers. (a) L^2: $\int (f-g)^2 dt$. (b) Integral wavefields method: $\int (\int f - \int g)^2 dt$. After data normalization, (c) NIM measures $\int (F-G)^2 dt$, while W_2 considers $\int (F^{-1}-G^{-1})^2 dt$ and W_1 considers $\int |F^{-1}-G^{-1}| dt$.

synthetic data in time, NIM first preconditions the data and then takes the integration. The objective function is:

$$J_3(m) = \frac{1}{2}\sum_r \int |Q(f(x_r,t;m)) - Q(g(x_r,t))|^p \, dt, \qquad (4)$$

where Q is transformation of the wavefield u, defined as:

$$Q(u)(x_r,t) = \frac{\int_0^t P(u)(x_r,\tau)d\tau}{\int_0^T P(u)(x_r,\tau)d\tau}. \qquad (5)$$

The operator P is included to make the data nonnegative. Three common choices are $P_1(u) = |u|$, $P_2(u) = u^2$ and $P_3 = E(u)$, which correspond to the absolute value, the square and the envelop of the signal (Liu et al., 2012).

Despite the fact that both methods are measuring the L^2 misfit, there are three different features in NIM compared with conventional FWI. Data sets are normalized to be nonnegative, mass balanced and integrated in time. The first two are exactly the prerequisite of optimal transport based misfit functions, i.e. the Wasserstein metrics.

Optimal transport

Optimal transport refers to the problem of seeking the minimum cost required to transport mass of one distribution into another given a cost function, e.g. $|x-y|^p$. The mathematical definition of the distance between the distributions $f : X \to \mathbb{R}^+$ and $g : Y \to \mathbb{R}^+$ can then be formulated as

$$W_p^p(f,g) = \inf_{T_{f,g}\in\mathscr{M}} \int_X |x - T_{f,g}(x)|^p f(x)\,dx \qquad (6)$$

where $\mathscr{M}$ is the set of all maps $T_{f,g}$ that rearrange the distribution f into g (Villani, 2003).

The optimal transport formulation requires non-negative distributions and equal total masses, $\int f(x)dx = \int g(x)dx$, which are not natural for seismic signals. Therefore a proper data normalization is required before inversion. Datasets f and g can be rescaled to be nonnegative with values in range $[0,1]$, and to have equal mass. This step is exactly the same as the one in Equation (5) in NIM.

We can compare the data trace by trace and use the Wasserstein metric (W_p) in 1D to measure the misfit. The overall misfit is then

$$J_4(m) = \sum_{r=1}^{R} W_p^p(f(x_r,t;m), g(x_r,t)), \qquad (7)$$

where R is the total number of traces. In this paper, we mainly discuss the quadratic Wasserstein metric (W_2) when $p=2$ in (6) and (7).

PROPERTIES

Next we discuss the similarities and difference among the misfit functions mentioned above. We will regard f and g as the synthetic and observed data from one single trace as an 1D problem.

Relations among misfit functions

Conventional full-waveform inversion measures the L^2 norm difference $\int |f(t)-g(t)|^2 dt$, indicated by the shaded part in Figure 1a. The integral wavefields misfit functional first integrates f and g in time, and then measures their L^2 misfit (3). The integral wavefields can be viewed as wavefields produced by a low-passed seismic wavelet. The created lower frequency components (in Figure 1b) can properly explain the improvement in inversion (Huang et al., 2014).

With a proper normalization method, it is possible to scale the data to have nonnegativity and mass balance. This step is essential for both NIM and W_2. Since processing data trace-by-trace is an 1D problem, we are able to solve the optimal transport problem exactly (Villani, 2003). The optimal map is the unique monotone rearrangement of the density f into g. In order to compute the quadratic Wasserstein metric, we need the cumulative distribution functions F and G and their inverses F^{-1} and G^{-1}. The explicit formulation for the 1D Wasserstein metric is:

$$W_p^p(f,g) = \int_0^1 |F^{-1}(x) - G^{-1}(x)|^p dx. \qquad (8)$$

The interesting fact is that W_2 computes the L^2 misfit between F^{-1} and G^{-1} (Figure 1c), while the objective function of NIM measures the L^2 misfit between F and G, i.e. $\int_0^T |F(t)-G(t)|^2 dt$ (Figure 1c). This is identical to the mathematical norm of Sobolev space H^{-1}, $||f-g||^2_{H^{-1}}$, given f and g are nonnegative and sharing equal mass.

Since F and G are both monotone increasing, one can show that there is an equivalency between NIM and W_2 misfit with

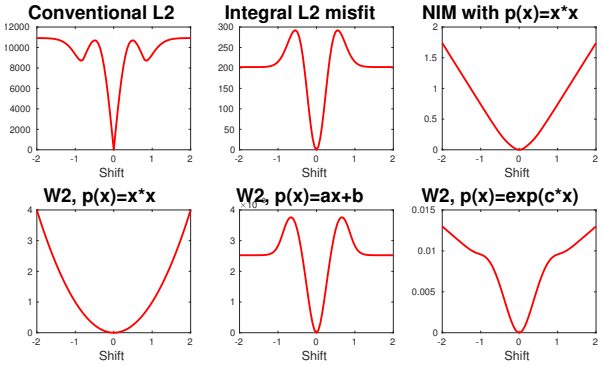


Figure 2: The misfit between $f(x)$ and $f(x-s)$ by six different misfit functions. First row shows conventional L^2 (left), intergral wavefield method (middle) and NIM with $p(x)=x^2$ (right). Second row shows the W_2 misfit with different normalization methods: $p(x)=x^2$ (left), $ax+b$ (middle) and $exp(c*x)$ (right).

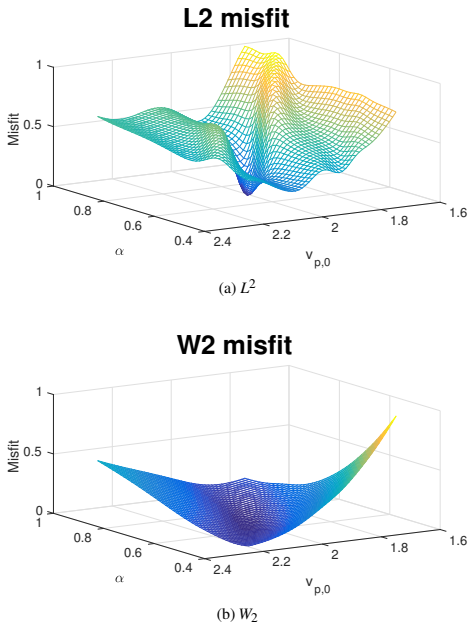


Figure 3: (a) Convexity plot of conventional L^2 (b) Convexity plot of trace-by-trace W_2 with normalization $p_2(x)=ax+b$

the same data normalization. Another demonstration of the similarity between NIM and optimal transport based metrics comes when $p=1$ in (4) and (8). These two misfits are the same since $\int |F(t)-G(t)|dt = \int_0^1 |F^{-1}(x)-G^{-1}(x)|dx$.

Mathematical connection between H^{-1} norm and W_2 norm

Next we move into a general case that f and g are synthetic and observed data in higher dimensions, satisfying nonnegativity and conservation of mass. To compute the quadratic Wasserstein metric, we solve the following Monge-Ampère equation (Brenier, 1991)

$$\det(D^2u(x)) = f(x)/g(\nabla u(x)) \tag{9}$$

If f and g are close enough and $g=(1+\varepsilon h+O(\varepsilon^2))f$, where h has mean zero, we can linearize (9) and also derive an approximation of the quadratic Wasserstein metric between f and g (Villani, 2003, p126-p127):

$$W_2^2(f,g) \approx \int_{\mathbb{R}^n} |\nabla\phi(x)|^2 f(x)dx = ||f-g||^2_{H^{-1}(d\mu)}, \tag{10}$$

where $d\mu = f(x)dx$. In one word, the quadratic Wasserstein metric is a weighted H^{-1} norm.

Besides, the dynamical characterization of the Wasserstein metric proposed by Benamou-Brenier (Benamou and Brenier, 2000) gives insights to consider that H^{-1} and W_2 belongs to the same class of measures. One can refer to Dolbeault et al. (2009) and Cardaliaguet et al. (2012) for more theoretical details, and Papadakis et al. (2014) for computational schemes. Mathematically, the misfits computed by NIM and W_2 are close also in higher dimensions.

Convexity

In order to illustrate the convexity of different objective functions, we borrow an example from Engquist and Froese (2014) that compares the misfit between a Ricker wavelet f and its shift $f(x-s)$. One can refer to the blue and red curves in Figure 1a. Here we plot the data misfits as a function of s in Figure 2. Conventional L^2, Intergral L^2 and NIM are compared on the first row. The second row displays W_2 misfits with three different scaling functions.

The figure on the top left for the conventional L^2 is the motivation of (Engquist and Froese, 2014) to bring the quadratic Wasserstein metric into seismic inversion. Such many local minima in the figure are not in favor of gradient-based optimization. The graph on the top middle is result of integral wavefields misfit functional. It creates lower frequency component, which decrease the chance of cycle skipping. Although having less local minima than conventional L^2, this method is still ill-posed in inversion. Integrating the wavefields or integrating the source may help invert the low wavenumber component of velocity, but still suffers from cycle skipping issues.

As demonstrated by Engquist et al. (2016), the squared Wasserstein metric has several properties that make it attractive as a choice of misfit function. One highly desirable feature is its convexity with respect to several parameterizations. However, the convexity highly depends on the data normalization method to satisfy nonnegativity and mass balance. The curves in the second row of Figure 2 are W_2^2 distance with different scaling functions: $p_1(x)=x^2$, and $p_2(x)=ax+b$ and $p_3(x)=\exp(c\cdot x)$. Theoretically p_1 gives perfect convexity, but having difficulty in inversion with adjoint-state method. From Taylor expansion p_3 is very close to p_2 when c is small, but easy to blow up with large c. Our current choice is to normalize data with p_2, but it is worth thinking a new normalization function that is able to preserve the convexity better.

It is interesting to compare the graph for NIM (upper right)

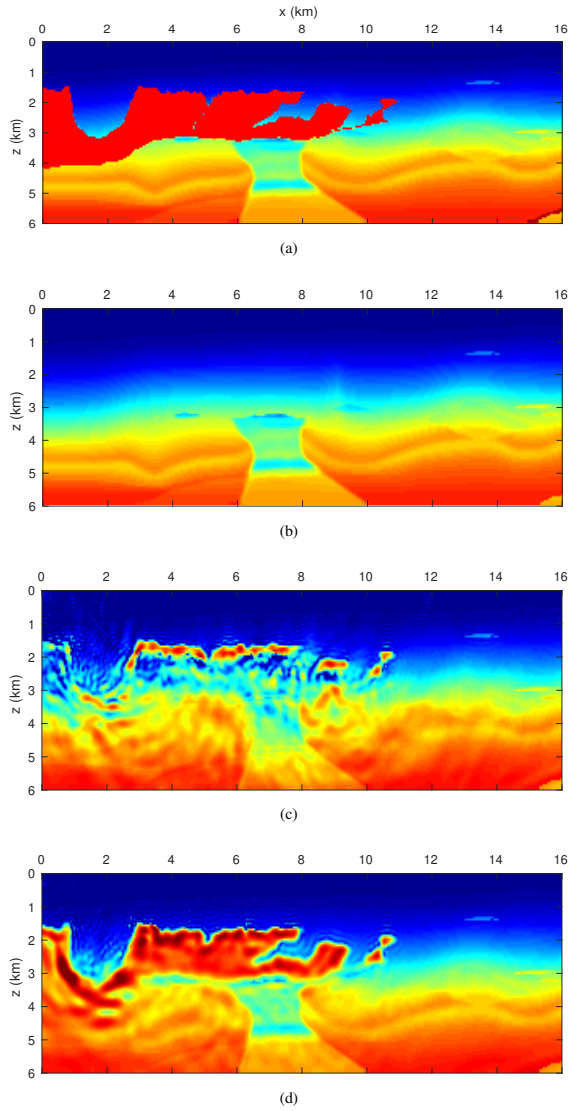


Figure 4: (a) True model velocity (b) Initial velocity (c) Inversion result using L^2 (d) Inversion result using W_2

with the one of W_2 (lower left) both of which are using the same normalization function (p_1) and globally convex with respect to the shift s. When $f(x)$ and $f(x-s)$ are close (i.e. $|s|$ is small), W_2 is a weighted H^{-1} as (10) states. Both curves have good convexity as $O(s^2)$ around zero. As $|s|$ gets larger, $W_2^2(f, f_s)$ is still $O(s^2)$, while the misfit measured by NIM is $O(s)$. The convexity of NIM becomes a bit weaker.

Finally we present a convexity result in model domain. We borrow the example from Métivier et al. (2016). The velocity model is assumed to vary linearly in depth as $v(x,z) = v_{p,0} + \alpha z$, where v_0 is the starting velocity on the surface, α is vertical gradient and z is depth. The reference for $(v_{p,0}, \alpha)$ is $(2km/s,\ 0.7s^{-1})$, and we plot the misfit curves with $\alpha \in [0.4, 1]$ and $v_0 \in [1.75,\ 2.25]$ on 41×45 grid in Figure 3. We observe many local minima and maxima in Figure 3a. Although W_2 is not convex in data domain with normalization method $p_2(x) = ax + b$ (Figure 2), the curve for W_2 (Figure 3b) is globally convex in model parameters $v_{p,0}$ and α. It demonstrates the capacity of W_2 in mitigating cycle skipping issues.

NUMERICAL EXAMPLE

In this section, we use a part of the BP 2004 benchmark velocity model (Billette and Brandsberg-Dahl, 2005) (Figure 4a) and an initial model without the upper salt part (Figure 4b) to do inversion with W_2 and L^2 norm respectively. A fixed-spread surface acquisition is used, involving 11 shots located every 1.6km on top. A Ricker wavelet centered on 5Hz is used to generate the synthetic data with a bandpass filter only keeping 3 to 9Hz components. We stopped the inversion after 300 L-BFGS iterations.

Here we precondition the data with function $p_2(x) = ax + b$ to satisfy the nonnegativity and mass balance in optimal transport. Inversion with trace-by-trace W_2 norm successfully construct the shape of the salt bodies (Figure 4d), while FWI with the conventional L^2 failed to recover boundaries of the salt bodies as shown by Figure 4c.

CONCLUSION

In this paper, we summarize and compare four misfit functions: the conventional least-squares inversion (L^2), the integral wavefields misfit function, the Normalized Integration Method (NIM), and the quadratic Wasserstein metric (W_2) from optimal transport. The L^2 norm is popular in general inverse problems, but suffers from cycle skipping in seismic inversion. The other three methods all incorporate the idea of integration the waveforms. Integration helps in enhancing the lower frequency component, but cannot avoid local minima coming from the oscillatory periodicity of the data. It is ideal to have a preconditioning operator which can "break" the periodicity and "record" the previous data information in time.

One solution is to combine the nonnegativity and integration in time together. Both NIM and the quadratic Wasserstein metric include these ideas as essential steps. A detailed discussion illustrates that the quadratic Wasserstein metric and the H^{-1} norm which NIM computes belong to the same family of mathematical measures. Moreover, H^{-1} and W_2 become equivalent when the two data sets are close. The analysis among these misfit functions of FWI brings additional insights into the importance of seismic data preconditioning, which also can be seen in examples of large scale FWI.

ACKNOWLEDGMENTS

We thank Sergey Fomel, Junzhe Sun and Zhiguang Xue 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.

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2017 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Alford, R., K. Kelly, and D.M. Boore, 1974, Accuracy of finite-difference modeling of the acoustic wave equation: Geophysics, **39**, 834–842, http://dx.doi.org/10.1190/1.1440470.

Benamou, J.-D., and Y. Brenier, 2000, A computational fluid mechanics solution to the monge-kantorovich mass transfer problem: Numerische Mathematik, **84**, 375–393, http://dx.doi.org/10.1007/s002110050002.

Billette, F., and S. Brandsberg-Dahl, 2005, The 2004 bp velocity benchmark: Presented at the 67th EAGE Conference & Exhibition.

Brenier, Y., 1991, Polar factorization and monotone rearrangement of vector-valued functions: Communications on pure and applied mathematics, **44**, 375–417, http://dx.doi.org/10.1002/(ISSN)1097-0312.

Brossier, R., S. Operto, and J. Virieux, 2010, Which data residual norm for robust elastic frequency-domain full waveform inversion?: Geophysics, **75**, no. 3, R37–R46, http://dx.doi.org/10.1190/1.3379323.

Cardaliaguet, P., G. Carlier, and B. Nazaret, 2012, Geodesics for a class of distances in the space of probability measures: Calculus of Variations and Partial Differential Equations, 1–26, https://doi.org/10.1007/s00526-012-0555-7.

Chauris, H., D. Donno, and H. Calandra, 2012, Velocity estimation with the normalized integration method: Presented at the 74th EAGE Conference and Exhibition incorporating EUROPEC 2012, https://doi.org/10.3997/2214-4609.20148721.

Dolbeault, J., B. Nazaret, and G. Savaré, 2009, A new class of transport distances between measures: Calculus of Variations and Partial Differential Equations, **34**, 193–231, http://dx.doi.org/10.1007/s00526-008-0182-5.

Donno, D., H. Chauris, and H. Calandra, 2013, Estimating the background velocity model with the normalized integration method: 75th Annual International Conference and Exhibition incorporating SPE EUROPEC, EAGE, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.20130411.

Engquist, B., and B. D. Froese, 2014, Application of the Wasserstein metric to seismic signals: Communications in Mathematical Sciences 12, 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, http://dx.doi.org/10.4310/CMS.2016.v14.n8.a9.

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, http://dx.doi.org/10.1190/1.3112572.

Huang, G., H. Wang, and H. Ren, 2014, Two new gradient precondition schemes for full waveform inversion: arXiv preprint arXiv:1406.1864.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering: Theory and application: Society for Industrial and Applied Mathematics, 206–220.

Liu, J., H. Chauris, and H. Calandra, 2012, The normalized integration method-an alternative to full waveform inversion?: Presented at the 25th Symposium on the Application of Geophysics to Engineering & Environmental Problems, https://doi.org/10.3997/2214-4609.20144373.

Luo, J., and R.-S. Wu, 2015, Seismic envelope inversion: Reduction of local minima and noise resistance: Geophysical Prospecting, **63**, 597–614, http://dx.doi.org/10.1111/1365-2478.12208.
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, http://dx.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, http://dx.doi.org/10.1088/0266-5611/32/11/115008.
Papadakis, N., G. Peyré, and E. Oudet, 2014, Optimal transport with proximal splitting: SIAM Journal on Imaging Sciences, **7**, 212–238, http://dx.doi.org/10.1137/130920058.
Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Reviews of Geophysics, **20**, 219–232, http://dx.doi.org/10.1029/RG020i002p00219.
Villani, C., 2003, Topics in optimal transportation: American Mathematical Society, Graduate Studies in Mathematics 58.
Virieux, J., A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, 2017, In 6. An introduction to full waveform inversion: Encyclopedia of Exploration Geophysics, R1-1–R1-40, http://dx.doi.org/10.1190/1.9781560803027.entry6.
Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, 74, no. 6, WCC1–WCC26, http://dx.doi.org/10.1190/1.3238367.
Warner, M., and L. Guasch, 2014, In Adaptive waveform inversion: Theory: 84th Annual International Meeting, SEG, Expanded Abstracts, 1089–1093, http://dx.doi.org/10.1190/segam2014-0371.1.
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 preprint arXiv:1612.05075.
Zhu, H., and S. Fomel, 2016, Building good starting models for full-waveform inversion using adaptive matching filtering misfit: Geophysics, 81, no. 5, U61–U72, http://dx.doi.org/10.1190/geo2015-0596.1