The application of an optimal transport to a preconditioned data matching function for robust waveform inversion

Bingbing Sun and Tariq Alkhalifah, King Abdullah University of Science and Technology

SUMMARY

Full Waveform Inversion updates the subsurface model iteratively by minimizing a misfit function, which measures the difference between observed and predicted data. The conventional l_2 norm misfit function is widely used as it provides a simple, sample by sample, high resolution misfit function. However it is susceptible to local minima if the low wavenumber components of the initial model are not accurate. A deconvolution of the predicted and observed data offers an extend space comparison, which is more global. The matching filter calculated from the deconvolution has energy focussed at zero lag, like a Dirac Delta function, when the predicted data matches the observed ones. We use the Wasserstein distance to measure the difference between the matching filter and a Dirac Delta function. Unlike data, the matching filter can be easily transformed to a distribution satisfying the requirement of optimal transport theory. Compared with the conventional normalized penalty applied to non-zero lag energy in the matching filter, the new misfit function is a metric and has solid mathematical foundation based on optimal transport theory. Both synthetic and real data examples verified the effectiveness of the proposed misfit function.

INTRODUCTION

Full waveform inversion (Virieux and Operto, 2009) is a nonlinear inversion process, and thus, we iteratively update the subsurface model to reduce the mismatch between the predicted and observed seismic data. Mathematically, we design a misfit function that characterizes such mismatch. Designing a misfit function is an important ingredient of the optimization problem: a well-behaved misfit function would release the requirement for a good initial velocity model or usable low frequency signal in the data and resolve the so-called ”cycle skipping” issue. Over the past decade, the least square l_2 norm was widely used as a misfit function for its simplicity and its potential for high resolution models, but it surfers from ”cycle skipping”. Recently, new and more advanced misfit functions were proposed, such as matching filter based misfit function (Van Leeuwen and Mulder, 2008, 2010; Luo and Sava, 2011; Warner and Guasch, 2016) and optimal transport misfit function (Engquist and Froese, 2014; Engquist et al., 2016; Mtivier et al., 2016; Yang et al., 2018; Yang and Engquist, 2018; Qiu et al., 2017). Those newly proposed methods transforms the local, sample by sample comparison, to a global one, trace by trace, or even traces by traces. As expected, the resulting misfit function shows more convex behavior and can reasonably mitigate the ”cycle skipping” issue.

A matching filter is computed for each trace by deconvolving the computed data from the observed data (Luo and Sava, 2011). If the velocity is correct, the resulting matching filter should resemble Dirac Delta function with energy focussed at zero time lag. We can design a misfit function by applying a penalty on the time lag (Luo and Sava, 2011; Huang et al., 2017) or with an additional normalization term to gain better convexity (Warner and Guasch, 2016). In principle, all those approaches try to measure the departure of the matching filter from a Dirac Delta function.

In this abstract, we propose a new misfit function by combing the matching filter measurement with optimal transport theory resulting in a more elegant way of measuring the distance between the matching filter and the Dirac Delta function. Current implementations of optimal transport in FWI is limited to measuring the distance between the predicted and observed data directly (Engquist and Froese, 2014; Mtivier et al., 2016; Yang et al., 2018). However, the optimal transport measurement using Wasserstein distance requires the compared variables be distributions, i.e., they should be positive and their integration equals 1. As seismic signals are oscillatory, they do not meet such a criterion. However, transforming the seismic signal into a distribution (Qiu et al., 2017) directly would either alter its amplitude or phase, which would potentially make the subsequent inversion unstable and possibly inaccurate. In order to resolve this issue, instead of measuring the distance between the predicted and computed data directly, we suggest computing a matching filter between the predicted and computed data and then measure the distance between the resulting matching filter and the Dirac Delta function. A precondition would transform the resulting matching filter to a distribution and since we are not modifying the predicted or observed data directly, the phase and amplitude of the seismic signal are preserved and the following inversion process should be stable and accurate. Compared with previous approaches for measuring the distance such as using a penalty method, the new misfit function is a metric and has solid mathematical foundation based on the optimal transport theory.

In the following, we review of the the optimal transport theory and then propose a misfit function by combing the matching filter and optimal transport theory. Finally, we use a modified Marmousi model and a marine real dataset from Australia to demonstrate the effectiveness of the proposed method.

CONVENTIONAL OPTIMAL TRANSPORT: WASSERSTEIN DISTANCE BETWEEN PREDICTED AND OBSERVED DATA

There are several types of misfit functions designed based on optimal transport theory (Engquist et al., 2016; Mtivier et al., 2016; Yang et al., 2018). Here, we mainly follow the approach of Yang et al. (2018). We start with a review of the method and for details, we refer you to Yang et al. (2018). Considering the predicted data $p(t)$ for the current available model and observed data $d(t)$, the optimal transport suggests that we can use the Wasserstein distance to design a misfit function as follows:

$$J = \min_T \int |t - T(t)|^2 p(t) dt, \quad (1)$$

where T is the transport plan, which maps the mass in distribution of p into d. For a 1D problem, an explicit formula exits for the Wasserstein distance :

$$J = \int |t - D^{-1}(P(t))|^2 p(t) dt, \quad (2)$$

where D^{-1} is the inverse function of D, D and P are the cumulative distribution functions: $D(t) = \int_0^t d(t')dt', P(t) = \int_0^t p(t')dt'$. The optimal transport theory requires that functions d and p be distributions: $d(t) \geq 0$, $p(t) \geq 0$ and $\int d(t)dt = \int p(t)dt = 1$. However, the seismic signal is oscillatory with zero mean. Preconditioning of the data is often needed to fulfill this requirement, e.g., adding a large value to make the signal nonnegative, followed by a normalization to make the summation equals 1 (Yang and Engquist, 2018). As we suggested before, for the conventional optimal transport approach, this modification of the predicted and observed data would alter the amplitude or phase of the seismic signal, which may negatively impact the inversion and cause it to be unstable and inaccurate.

A NEW MISFIT FUNCTION : WASSERSTEIN DISTANCE BETWEEN MATCHING FILTER AND DIRAC DELTA FUNCTION

Conventional optimal transport approaches measure the Wasserstein distance between the predicted data $p(t)$ and observed data $d(t)$ directly. In this section, we propose to measure Wasserstein distance between a matching filter extracted from deconvolving the observed data from the predicted data and Dirac Delta function instead. Thus, at first, given observed data $d(t)$ and computed data $p(t)$, we compute a matching filter $w(t)$:

$$d(t) * w(t) = p(t), \quad (3)$$

where $*$ denotes the convolution operation. Equation 3 is a linear equation, and the matching filter can be computed either in the time domain or in frequency domain as:

$$w(t) = \mathscr{F}^{-}\left[\frac{\mathscr{F}[p(t)]\overline{\mathscr{F}[d(t)]}}{\mathscr{F}[d(t)]\overline{\mathscr{F}[d(t)]} + \varepsilon}\right], \quad (4)$$

where $\mathscr{F}$ and $\mathscr{F}^{-}$ denote the Fourier transform and its inverse, respectively, overline denotes for the complex conjugate, and ε has a small positive value to avoid dividing by zero. Next, we measure the Wasserstein distance between the matching filter $w(t)$ and the Dirac Delta function. In order to full-fill the requirement of the optimal transport theory, we precondition and modify $w(t)$ to be a distribution. We suggest to square it and normalize it as follows:

$$w'(t) = \frac{w^2(t)}{\int w^2(t)dt} = \frac{w^2}{||w||_2^2}. \quad (5)$$

When the model parameters are accurate, the resulting matching filter reduces to a "Dirac Delta function", this means the "Dirac Delta function" $\delta(t)$ is the target. Based on the theory of optimal transport, we use equation 2 directly and the new misfit function becomes

$$J(w', \delta(t)) = \frac{\int |t - \Delta^{-1}(W'(t))|^2 w^2(t) dt}{||w||_2^2}. \quad (6)$$

Here, we use Δ and $W'(t)$ to denote the commutative distribution function for Dirac Delta function $\delta^2(t)$ and the normalized matching filter $w'(t)$ respectively. Δ^{-1} is the inverse function for the commutative distribution function Δ. Because of the singularity involved in Dirac Delta function, in practice, we use a Gaussian function with a small standard deviation to approximate Dirac Delta function. For the proposed misfit function, in discrete form, the adjoint source can be computed as:

$$\delta s = \left(\frac{\partial w}{\partial p}\right)^H \left(\frac{\partial w'}{\partial w}\right)^H \left(\frac{\partial J}{\partial w'}\right)^H, \quad (7)$$

where H denotes complex transpose, the term $\left(\frac{\partial J}{\partial w'}\right)^H$ is the same as adjoint source for the conventional optimal transport misfit function (Yang and Engquist, 2018):

$$\left(\frac{\partial J}{\partial w'}\right)^H = \left[U \text{diag}\left(\frac{-2w'(t)}{\delta(\Delta^{-1}(W'(t)))}\right) + \text{diag}(t - \Delta^{-1}(W'(t))\right] (t - \Delta^{-1}(W'(t))), \quad (8)$$

where U is an upper triangular matrix whose nonzero values are 1. The term $\left(\frac{\partial w'}{\partial w}\right)^H$ can be computed using equation 5:

$$\left(\frac{\partial w'}{\partial w}\right)^H = 2\left[\text{diag}(w(t)) - w'(t)(w(t))^H\right). \quad (9)$$

Term $\left(\frac{\partial w}{\partial p}\right)^H$ can be obtained from the adjoint of equation 4:

$$\left(\frac{\partial w}{\partial p}\right)^H = \mathscr{F}^{-}\text{diag}\left[\frac{\mathscr{F}[d(t)]}{\mathscr{F}[d(t)]\overline{\mathscr{F}[d(t)]} + \varepsilon}\right]\mathscr{F} \quad (10)$$

COMPARISON WITH AWI

Transforming the time coordinates so that the zero lag is at 0.5, and the time axis scaled between 0 and 1, the adaptive waveform inversion (AWI) misfit function (Warner and Guasch, 2016) can be expressed as:

$$J_{\text{AWI}} = \frac{\int |t - 0.5|^2 w^2(t) dt}{||w||_2^2} \quad (11)$$

If we look back to the original definition of the optimal transport in equation (1) where $T(t)$ would define the mass transport plan from distribution $w'(t)$ to a Delta distribution. As the Delta function has only energy at time $t = 0.5$, the transport plan is fixed to be $T(t) = 0.5$. In this cases, our misfit function would reduce to AWI exactly by setting the transport plan in equation (6) to be 0.5, i.e., $\Delta^{-1}(W'(t)) = 0.5$. However, for real applications, in which the seismic signal is band limited, we could use band limited delta function rather than a singular one. In this case, $\Delta^{-1}(W'(t))$ provides the best transport plan for the given approximated target Delta function. In short, our objective function is derived from optimal transport theory, and thus, has a more general form. On the other hand, AWI was derived from differential semblance analysis (DSO) concept, combined with a normalization factor that helped convergence. Well, our objective function demonstrates that the normalization factor is necessary. Figures 1a and b show the role of $\Delta^{-1}(W'(t))$ for different choices of near zero lag functions as objectives for optimal transport, compared to setting it to 0.5 as in AWI. In both figures, we use a matching filter given by a random function. We can see from Figure 1a that when we approximate the Dirac Delta function with a Gaussian distribution of relatively large standard deviation, the resulting penalty would increase, followed by a decrease, while AWI continues to increase (equals to $(t-0.5)^2$). When we make the Gaussian thin (low standard deviation of the distribution) and it becomes more singular as shown in Figure 1b, as we discussed before it would reduce to AWI with almost the same penalty.

EXAMPLES

In this section, we apply our approach to invert for the modified Marmousi model. The true velocity v_{true} shown in Figure 2a extends 2 km in depth and 8 km, laterally. The initial velocity is shown in Figure 2b. The dataset is modeled using 80 shots with a source interval 100 m and 400 receivers with an interval of 20 m. We use a free surface boundary condition in the modeling and this adds more nonlinearity and makes the dataset more challenging for inversion . The source wavelet is a Ricker wavelet with a 10 Hz peak frequency. We mute the data below 3 Hz to verify that our proposed method is capable of overcoming the cycle skipping for data free of low frequency. We invert the dataset in the time domain using the proposed misfit function and the conventional l_2 norm misfit function, with the highest frequency equal to 5 Hz,10 Hz and 20 Hz, sequentially. The final inverted result for the l_2 and proposed method are shown in Figures 2c and 2d, respectively. From the result, we can see that the proposed method better recovers the true model both in the shallow and deeper parts compared to the l_2 norm conventional FWI, as l_2 norm fails in many areas dues to cycle skipping especially at depth.

The second example is a marine real data set from offshore Australia (Sun and Alkhalifah, 2018). The offset range is from 160 to 8200 m. The initial velocity model converted form RMS velocity is given in Figure 3a. We perform the inversion using the proposed misfit function with a low pass filter applied to the data equal to 3 Hz, 5 Hz, 10 Hz, 20 Hz, 40 Hz, sequentially. During the inversion, TV regularization is used to reduce the noise. The inverted model is shown in Figure 3b. The updated model shows consistent structures and high resolution due to the high frequency inversion and TV regularization. In the left panels of Figures 4a and 4b, we show one selected common shot gather from the initial and inverted models. We compare it with the recorded shot gather at the same location in the right panel. Clearly, the observed model reproduces the data that batter matches the observed data, especially at the larger offsets where cycle skipping usually happens, and it is evident for the initial model. Considering the initial velocity model is obtained from a crude RMS velocity, we attribute the reasonable good result to the proposed misfit function ability to handle cycle skipped data.

CONCLUSION

We proposed a misfit function, which utilizes a matching filter between the observed and predicted data, and uses the optimal transport concept to build a model that transforms the matching filter to a form that makes the predicted data fit the observed one, and that is a Dirac Delta function. The matching filter, unlike data, is better suited to be transformed to a distribution, which is required for the optimal transport. The resulting objective function derived using the Wasserstein distance provides a much wider basin of attraction than newly developed objective functions dedicated to avoiding the cycle skip problem. A Marmousi synthetic and an offshore real data examples verified the effectiveness of the proposed method in resolving cycle skipping problem.

ACKNOWLEDGMENTS

We thank KAUST for its support and the Shaheen super computing laboratory for the computational recourses . We acknowledge CGG for providing the real data.

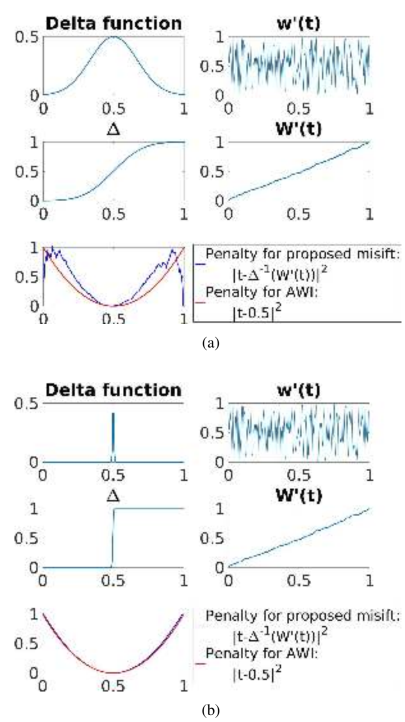
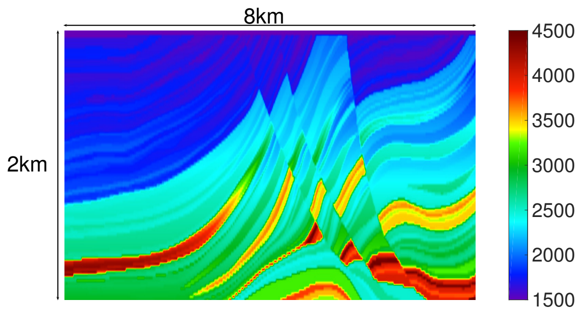
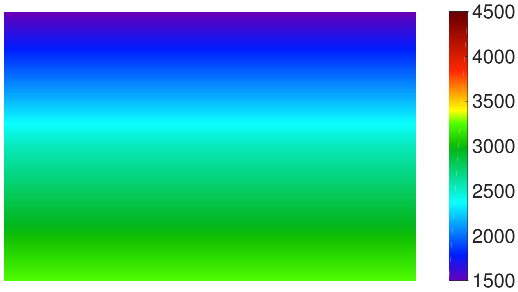


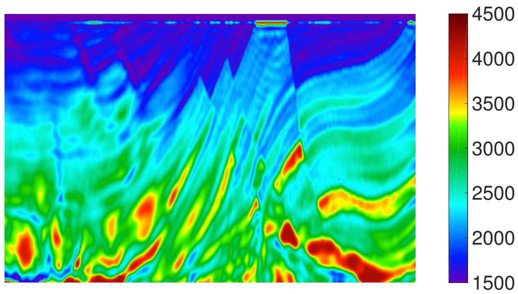
Figure 1: A comparison between the penalty term of our proposed misfit function and AWI when set the approximated Dirac Delta function to be a Gaussian function with (a) a large; and (b) a small standard deviation.



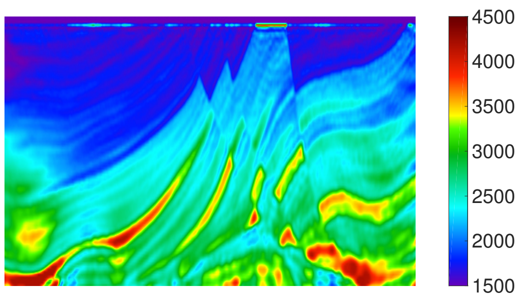
(a)



(b)

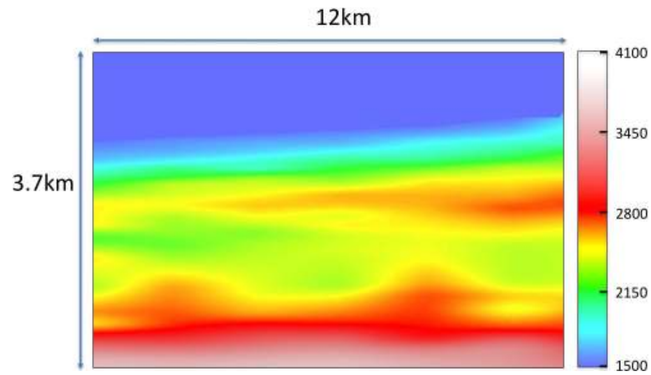


(c)

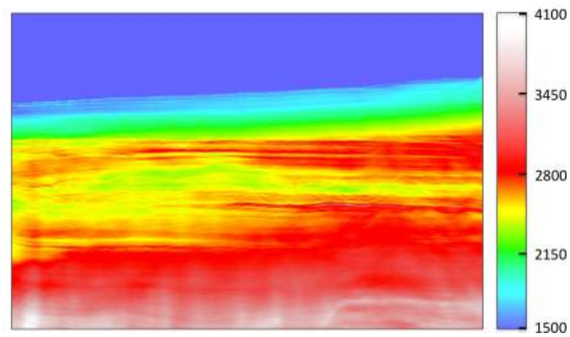


(d)

Figure 2: a) The true Marmousi velocity; b) the initial velocity; the inverted model based on c) the l_2 norm misfit function ; (d) the proposed misfit function.

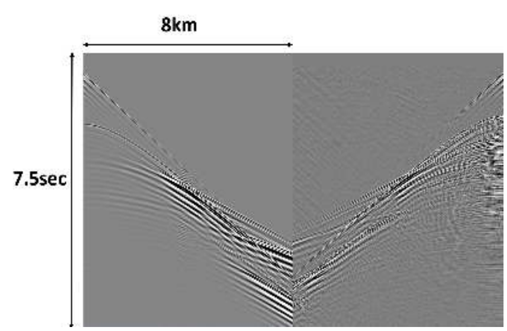


(a)

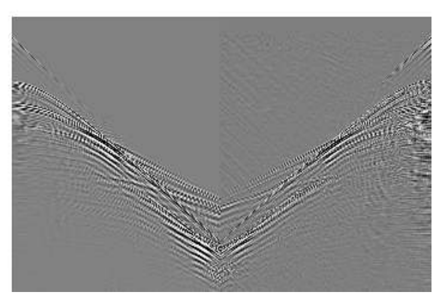


(b)

Figure 3: a) The initial model ; b) The inverted model.



(a)



(b)

Figure 4: Right panel is the shot record of the real seismic data while left panel is the modeled one from a) the initial model ; b) the inverted model.

10.1190/segam2018-2995285.1

REFERENCES

Engquist, B., and B. 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.

Huang, G., R. Nammour, and W. Symes, 2017, Full-waveform inversion via source-receiver extension: Geophysics, **82**, no. 3, R153–R171, https://doi.org/10.1190/geo2016-0301.1.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792, https://doi.org/10.1190/1.3627773.

Mtivier, L., R. Brossier, Q. Mrigot, 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.

Qiu, L., J. Ramos-Martnez, 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, https://doi.org/10.1190/segam2017-17681930.1.

Sun, B., and T. Alkhalifah, 2018, Automatic wave-equation migration velocity analysis by focusing subsurface virtual sources: Geophysics, **83**, U1–U8, https://doi.org/10.1190/geo2017-0213.1.

Van Leeuwen, T., and W. Mulder, 2008, Velocity analysis based on data correlation: Geophysical Prospecting, **56**, 791–803, https://doi.org/10.1111/gpr.2008.56.issue-6.

Van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation traveltime tomography: Geophysical Journal International, **182**, 1383–1394, https://doi.org/10.1111/j.1365-246X.2010.04681.x.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, https://doi.org/10.1190/1.3238367.

Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: Geophysics, **81**, no. 6, R429–R445, https://doi.org/10.1190/geo2015-0387.1.

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

10.1190/segam2018-2995285.1