Stereo optimal transport of the matching filter

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

SUMMARY

A recently introduced optimal transport of the matching filter (OTMF) provided us with a robust misfit function for reducing cycle skipping in Full-Waveform Inversion (FWI). Unlike the conventional L2-norm approach, OTMF computes a matching filter first by deconvolution of the predicted data with the measured ones and constructs the misfit function by measuring the Wasserstein distance W_2 between the resulting preconditioned matching filter and the target, i.e., the Dirac delta function. Compared to the conventional application of the optimal transport (OT) misfit in the data domain directly, OTMF applies the OT to the resulting matching filter avoiding the modification of the amplitude or the phase of the original seismic data. Measuring the distance between the resulting matching filter and a Dirac delta function using Wasserstein metric W_2 suggests a convex misfit function with respect to the time shifted signal. We propose a misfit function by stereo optimal transport of the matching filter (SOTMF), which takes the space coherency of the resulting matching filter into consideration. Compared to OTMF, SOTMF has an extra regularization term, which controls the variations of the resulting matching filter along the space (offset) axis, and the regularization term is formulated by Wasserstein distance between the matching filters of the nearby traces. Thus, in the framework of OT, SOTMF tries to focus the resulting matching filter to be a Dirac delta function in time and enhance its space coherency as well. We use the Marmousi example to show that SOTMF can reduce the cycle-skipping, and at the same time its result shows less artifacts than OTMF. A result using an anisotropic version of SOTMF waveform inversion on a real dataset also demonstrates the good performance of the approach.

INTRODUCTION

Full-waveform inversion (Tarantola, 1984) aims to produce high-resolution subsurface models that are able to produce data that fits the measured seismic waveforms. It is a nonlinear inversion process, and thus, we iteratively update the subsurface model to reduce the mismatch between the measured and the predicted seismic data. Mathematically, we design a misfit function to characterize such mismatch. Designing a misfit function that tolerates the expected poor quality and the limited band nature of field data is highly important. A well-behaved misfit function also would relax the requirement for a good initial velocity model and usable low frequency signals in the data, and thus, resolve the so-called cycle skipping issue.

Over the past decade, the least-square L2 norm is widely used as a misfit function for its simplicity and high resolution potential. It, however, suffers from the well known cycle skipping problem. Currently, new and more advanced misfit functions were proposed, those include, but not limited to, the matching filter based misfit function (Van Leeuwen and Mulder, 2008, 2010; Luo and Sava, 2011; Warner and Guasch, 2016; Huang et al., 2017; Sun and Alkhalifah, 2019), the optimal transport misfit function (Métivier et al., 2016; Yang et al., 2018; Yang and Engquist, 2018; Sun and Alkhalifah, 2018a), the instantaneous travel-time misfit function (Alkhalifah and Choi, 2014; Sun and Alkhalifah, 2018c) etc. Unlike the conventional L2-norm misfit function, those newly proposed methods transform the search space from local, sample by sample comparison, to global, trace by trace, comparisons.

A matching filter is computed for each trace by deconvolving the predicted data from the measured data (Luo and Sava, 2011). Sun and Alkhalifah (2019) proposed the optimal transport of the matching filter (OTMF) misfit function by using the matching filter in the optimal transport theory resulting in a more elegant way of introducing optimal transport in FWI. Current implementations of the optimal transport in FWI is limited to measuring the distance between the predicted and measured data directly (Engquist and Froese, 2014; Métivier 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 the measured data directly, OTMF suggests measuring the distance between a matching filter of the predicted and the measured data and the Dirac delta function. A precondition would transform the resulting matching filter which holds the data difference information to a distribution. Compared with previous approaches for measuring the distance such as using a penalty method, the new misfit function is a metric and has a solid mathematical foundation based on the optimal transport theory. In fact, Sun and Alkhalifah (2018a) show that adaptive waveform inversion (AWI) is a special case of OTMF misfit and the critical normalization term in AWI can be explained as a requirement for a probability distribution.

Current application of the OTMF method tries to focus the resulting matching filter to zero lag along the time axis. In this study, we propose to formulate a misfit function by stereo optimal transport of the matching filter (SOTMF). SOTMF includes a regularization term which controls the variation of the resulting matching filter along the space (offset) direction. Due to the lateral heterogeneity, such variations along the space (offset) direction can be strong, and the introduced regularization term would stabilize the inversion process. Differential Semblance Optimization (Gao et al., 2014) would suggest a space derivative of the resulting matching filter to implement such regularization; here, we use Wasserstein distance based on optimal transport theory to measure the distance between the matching filter distributions of nearby traces. The benefits of using Wasserstein distance for comparing the matching filter has been demonstrated (Sun and Alkhalifah, 2018a, 2019): The comparison would be more robust and less sensitive to the amplitude irregularity, and more importantly it preserves the convexity properties of the Wasserstein distance with regard to the time shift.

OPTIMAL TRANSPORT OF THE MATCHING FILTER (OTMF)

Here, we briefly review the OTMF approach. Readers can refer to Sun and Alkhalifah (2018a) and Sun and Alkhalifah (2019) for detailed information such as the adjoint source computation and the comparison with adaptive waveform inversion.

Conventional optimal transport approaches measure the Wasserstein distance between the predicted data $p(t)$ and the measured data $d(t)$ directly. The OTMF approach measures Wasserstein distance between a matching filter extracted from deconvolving the measured from the predicted data and the Dirac delta function instead. Thus, at first, given the measured data $d(t)$ and the predicted data $p(t)$, we compute a matching filter $w(t)$:

$$d(t) * w(t) = p(t), \tag{1}$$

where $*$ denotes the convolution operation. 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:

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

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 Wasserstein metric W_2 to measure the distance between the resulting matching filter and the Dirac delta function:

$$J_{\text{OTMF}} = W_2^2(w'(t), \delta(t)). \tag{3}$$

As we compute the Wasserstein distance per trace, it is a 1D optimal transport problem and an explicit formula exists (Yang and Engquist, 2018):

$$J_{\text{OTMF}}(w', \delta(t)) = \frac{\int |t - \Delta^{-1}(W'(t))|^2 w^2(t)dt}{||w||_2^2}. \tag{4}$$

Here, we use Δ and $W'(t)$ to denote the commutative distribution function for Dirac delta function $\delta(t)$ 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 the Dirac delta function, in practice, we use a Gaussian function with a small standard deviation to approximate the Dirac delta function.

STEREO OPTIMAL TRANSPORT OF THE MATCHING FILTER (SOTMF)

OTMF tries to focus the resulting matching filter in time direction only. An improvement can be made by taking the space coherency of the resulting matching filter into consideration. For example, Differential Semblance Optimization (Mulder and ten Kroode, 2002; Gao et al., 2014) would suggest enhancing the semblance of the resulting matching filter along space (offset) direction and propose the following function for regularization:

$$R_{\text{DSO}} = \sum_{x_i} \left[\frac{\partial w(t,x)}{\partial x} \right]^2_{x=x_i}, \tag{5}$$

where, x_i denotes the trace at different space locations (offset). Based on our framework of the optimal transport matching filter, we can design a regularization term by measuring the Wasserstein distance of nearby matching filters:

$$R_{\text{OTMF}} = \sum_{x_i} W_2^2 \left(w'(t,x_i), w'(t,x_{i+1}) \right). \tag{6}$$

The new form of comparison between two nearby matching filters using an optimal transport measure can provide better performance than the conventional DSO formula of equation 5: Due to the Wasserstein distance special features in probability distribution comparisons, it is less sensitive to change in the amplitude and preserves convexity with regard to time shift.

Thus, the final misfit function for SOTMF is

$$J_{\text{SOTMF}} = \sum_{x_i} \left[W_2^2(w'(t,x_i), \delta(t)) + \lambda W_2^2 \left(w'(t,x_i), w'(t,x_{i+1}) \right) \right], \tag{7}$$

where λ is a weighting term, which balances the focusing in the time direction and the coherence in the space direction. As the label "stereo" included in its name, SOTMF leads to a more robust inversion, which focuses the resulting matching filters in the time direction while maintaining their coherency in space (offset) direction as well.

EXAMPLES

In this section, we apply our approach to invert for the modified Marmousi model. The true velocity v_{true} shown in Figure 1a extends 2 km in depth and 8 km, laterally. The initial velocity is shown in Figure 1b. The dataset is modeled using 80 shots with a source interval 100 m and 400 receivers with an interval of 20 m. 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. In the inversion, we do not perform frequency continuation. Instead, we band-pass the dataset with highest frequency 10 Hz and perform the inversion directly. We iterate overs 200 iterations, and the final inverted result for the L2 norm, OTMF and SOTMF misfits are shown in Figures 1c to 1e respectively. From the results, the L2-norm misfit fails in many areas due to cycle skipping especially at depth while SOTMF can mitigate the cycle-skipping and outperform the OTMF misfit with less artifacts (the areas denoted by the black arrows).

The second example is a marine real data set from offshore Australia (Sun and Alkhalifah, 2018b). The offset range is from 160 to 8200 m. The initial vertical velocity is converted from RMS velocity given in Figure 2a, and we set the anisotropic parameter $\varepsilon = 0$ shown in Figure 2b for the initial model. We perform the VTI-FWI inversion using the proposed SOTMF misfit function with a low-pass filter applied to the data from 4 Hz to 9 Hz every 1 Hz sequentially and refine the obtained result using the L2-norm misfit function at the end for higher resolution. The final inverted model for the vertical velocity and the epsilon is shown in Figures 2c and 2d, respectively. We can see that the inverted velocity model shows consistent structures. In the right panels of Figures 3a and 3b, we show one selected common shot gather for the initial and the inverted models. We compare them with the recorded shot gather at the same location in the right panel. Clearly, the inverted model reproduces the data that better matches the measured 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 good convergence to the proposed misfit function's ability to handle cycle skipped data. In Figures 3c and 3d, we compare the RTM image for the initial model and the inverted model. It is clear that after VTI-FWI updating of the model, the image shown in Figure 3d becomes better focused (note the area denoted by the blue arrows). The improved focusing of a V-shaped fault denoted by the yellow dotted line further demonstrates the high accuracy of the inverted model.

CONCLUSION

We proposed a misfit function, which utilizes a matching filter between the measured 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. We suggest to apply a space regularization term, which measures the Wasserstein distance between the nearby matching filters. A Marmousi synthetic and an offshore real data examples verified the effectiveness and robustness of the proposed stereo optimal transport of the matching filter approach.

ACKNOWLEDGMENTS

We thank the members of SWAG group for useful discussions and the Shaheen super computing laboratory for the computational recourses. We acknowledge CGG for providing the offshore Australia real dataset.

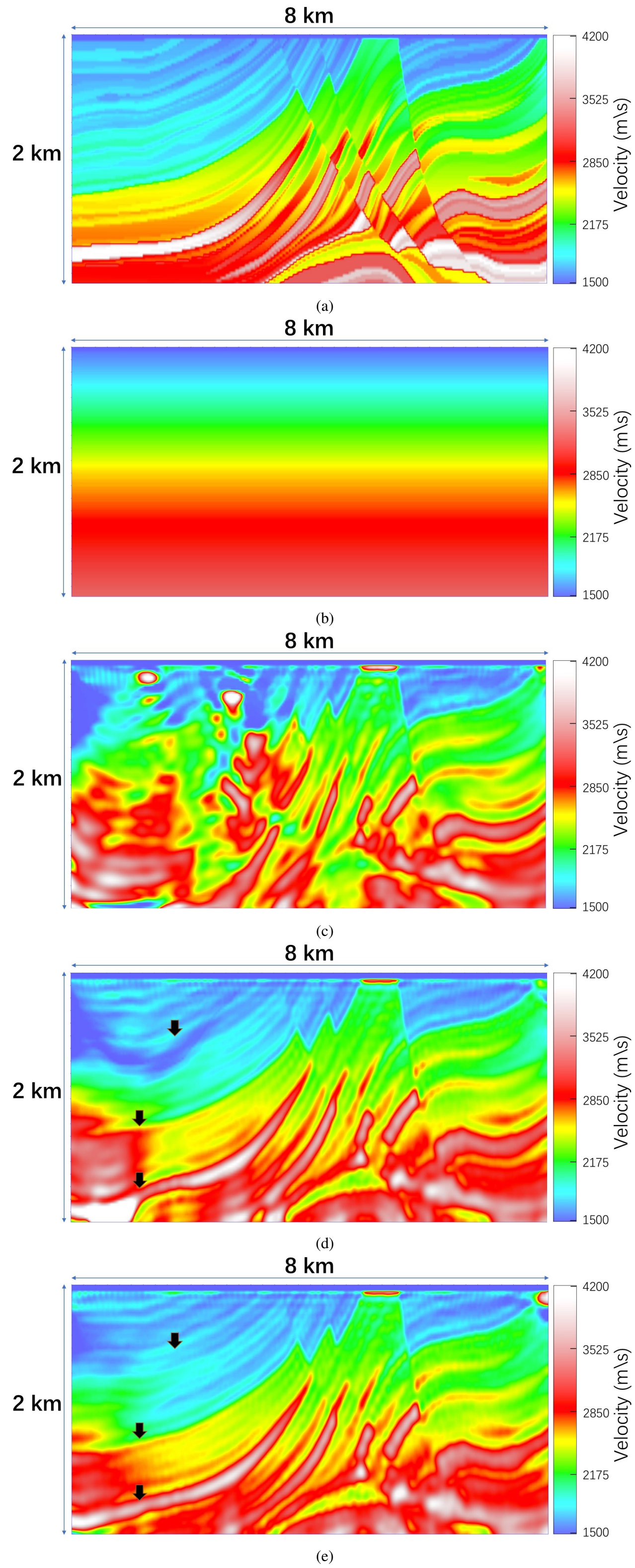

Figure 1: a) The true Marmousi velocity; b) the initial velocity; the inverted model based on c) the L2-norm misfit function; (d) the OTMF misfit; e) the SOTMF misfit.

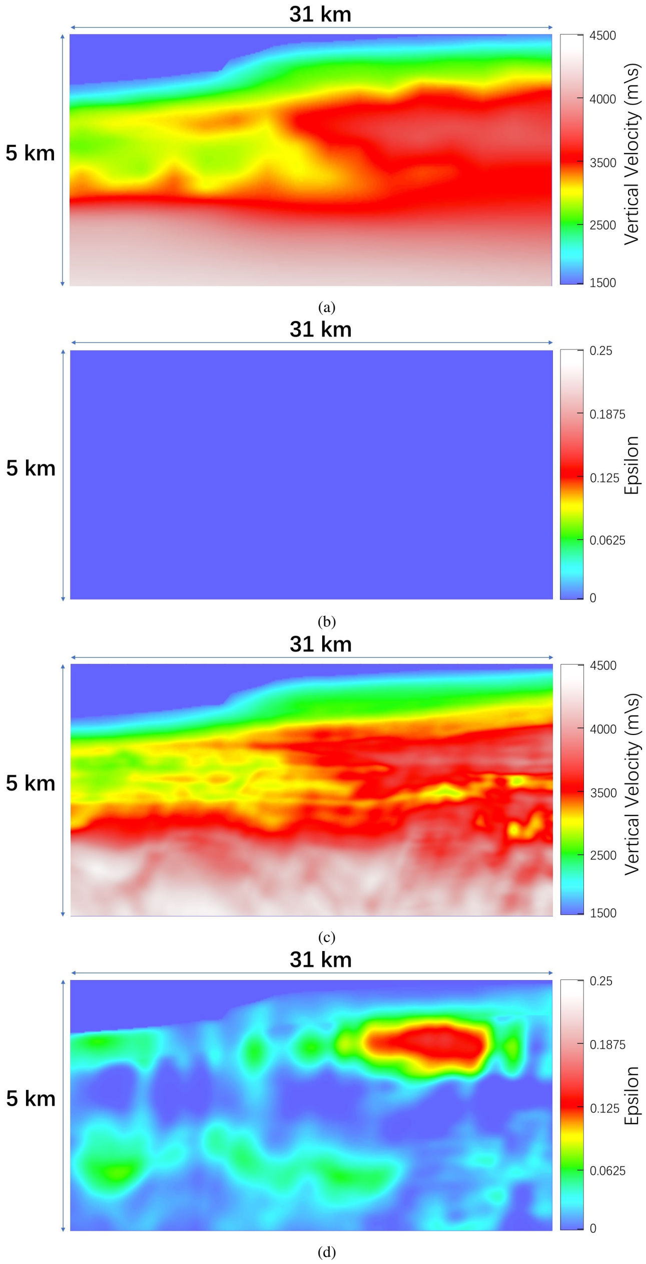

Figure 2: The initial model for a) the vertical velocity; b) the epsilon; the inverted model for c) the vertical velocity; d) the epsilon.

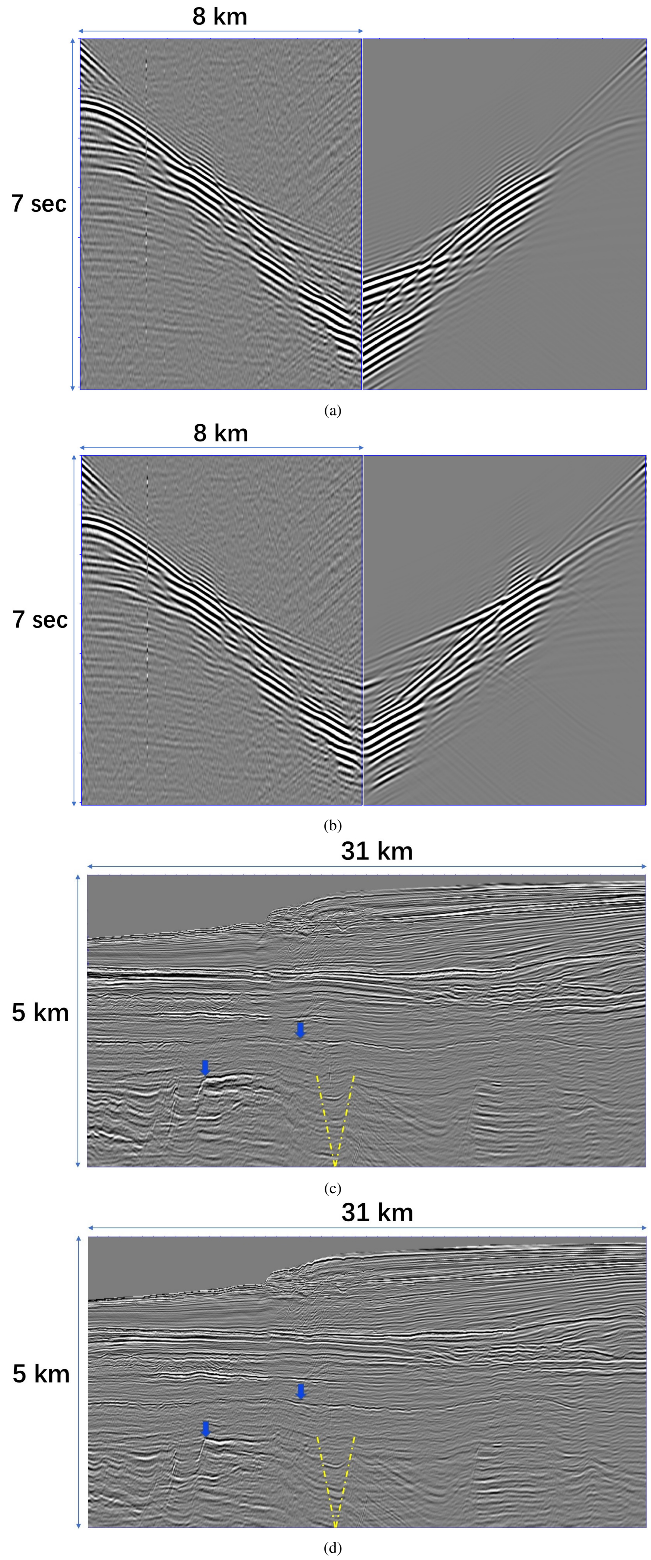

Figure 3: The record comparison of the measured data (left panel) and the modeled data by a) the initial model (right panel); b) the inverted model (right panel); the RTM image for c) the initial model; d) the inverted model.