2019 EAGE ANNUAL
81ST CONFERENCE + EXHIBITION
LONDON | THE UNITED KINGDOM

Tu_R08_02

Multidimensional Optimal Transport for 3D FWI: Demonstration on Field Data

J. Messud[1]*, A. Sedova[1]

[1] CGG

Summary

We discuss the advantages of multidimensional (in data space) optimal transport (OT) full waveform inversion (FWI). We show that a careful formulation leads to an enhanced coherency of the event continuity in the move-out direction and to an improved velocity update compared to the conventional, least squares based, cost function and also compared to the monodimensional (in data space) OT FWI. This is illustrated on both marine and land field datasets.

Introduction

Conventional full waveform inversion (FWI) is based on a least squares (LSQ) misfit function, which has proven to be effective for high-resolution velocity model building. However, LSQ FWI is plagued by local minima (due to cycle skipping), so that the local optimization process requires starting from a good initial model and, additionally, careful data processing is often required for successful field data applications. Many alternative misfit functions have been proposed to mitigate those issues. The common idea is to emphasize the kinematic information contained in the data, and to relieve sensitivity to the amplitude information. Among those, misfits based on the optimal transport (OT) theory, recently aroused attention in geophysics. OT FWI has the potential to reduce the local minima issue (Engquist et al., 2014). Encouraging results were obtained by Métivier et al. (2016a), Yang et al. (2018) and Poncet et al. (2018). In practice, OT implementation in FWI only modifies the back-propagated adjoint-source that can be viewed as the result of smart preprocessing of the LSQ residual, enhancing event coherency and low frequencies, thus the kinematic information (Métivier et al., 2016a; Yang et al., 2018; Poncet et al., 2018).

An unusual property of OT misfits is that they can be multi-dimensional in the data space (denoted by multiD in the following, not to be confused with the dimensionality of the velocity model). This means they can account for correlations between data samples in the time and receiver directions, which should result in an enhancement of the coherency of the events. In contrast, LSQ FWI is 0D in data space, meaning it considers each data sample independently and locally. Intermediately, 1D OT FWI accounts for correlations between data samples in the time direction but not in the receiver direction. The advantages of the multiD OT formulation are still to be further investigated: Yang et al. (2018) and Métivier et al. (2016a) investigated multiD OT v.s. 1D OT FWI, but their tests on synthetic data led to somewhat limited conclusions.

In this article, we demonstrate on field data the interest of multiD OT FWI. We show that a careful formulation leads to an increased event coherency along the move-out direction compared to 1D OT and LSQ FWI, producing an improved structural consistency in velocity updates. We illustrate how our implementation leads to better FWI results on both marine and land field datasets.

OT FWI implementations and multiD aspects

For each shot, we denote by $d_{obs}(\boldsymbol{x})$ the observed data and $d[\boldsymbol{m}](\boldsymbol{x})$ the data modelled using a subsurface model $\boldsymbol{m}$. The data space $\boldsymbol{X}$ is parameterized by the time and receiver positions. For a misfit measurement $\sum_{shots} J(d_{obs}, d[\boldsymbol{m}])$, the data-space gradient $\delta J/\delta d[\boldsymbol{m}]$ defines the adjoint-source, translated into the velocity-space gradient via the adjoint-state method. $c_p(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|_p$ denotes the L_p distance between vectors in the $\boldsymbol{X}$ space. The LSQ misfit and data-space gradient are

$$J_{L_2}(d_{obs}, d[\boldsymbol{m}]) = \frac{1}{2} c_2^2(d_{obs}, d[\boldsymbol{m}]) \quad \text{and} \quad \frac{\delta J_{L_2}}{\delta d[\boldsymbol{m}]} = \Delta d[\boldsymbol{m}] \qquad (1)$$

where $\Delta d[\boldsymbol{m}] = d_{obs}(\boldsymbol{x}) - d[\boldsymbol{m}]$ denotes the data residual.

There exist various formulations of OT applied to FWI, all related to Wasserstein distances. The p-Wasserstein distance for two probability densities d_1 and d_2 in the data space is

$$J_{W_p}^p(d_1, d_2) = \min_{\boldsymbol{T}} \int_{\boldsymbol{X}} c_p^p(\boldsymbol{x}, \boldsymbol{T}(\boldsymbol{x})) d_2(\boldsymbol{x}) d\boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{T} \in maps\ that\ rearrange\ d_2\ into\ d_1 \qquad (2)$$

where s.t. denotes "subject to constraint". Eq. (2) seeks the minimum cost to transport mass from d_2 to d_1, from the cost c_p^p point of view. An important aspect is that it requires positive d_1 and d_2 with equal masses, so that it cannot be readily applied to seismic data. To overcome this limitation, Yang et al. (2018) and Qiu et al. (2017) proposed positive transformations of the observed and modelled data followed by rescaling to the same mass. They chose the $p=2$ case, i.e. the squared 2-Wasserstein distance related to the LSQ cost c_2^2. Since solving eq. (2) in the multiD data space is computationally demanding, most of their applications consider a 1D data space, i.e. $\boldsymbol{X}$ is parameterized in time only with eq. (2) solved for each trace independently.

Métivier et al. (2016a) started from the 1-Wasserstein distance, i.e. $p=1$ in eq. (2) reformulated as a dual problem, that becomes linear. Then adding a bounding constraint allows for the use of the seismic data directly. This leads to the so-called Kantorovich-Rubinstein (KR) distance:

$$J_{KR}(d_{obs}, d[\boldsymbol{m}]) = \max_{\varphi} \int_{\boldsymbol{X}} \varphi(\boldsymbol{x}) \Delta d[\boldsymbol{m}](\boldsymbol{x}) d\boldsymbol{x} \quad \text{s.t. } |\varphi(\boldsymbol{x}) - \varphi(\boldsymbol{y})| \leq c_1(\boldsymbol{x}, \boldsymbol{y}) \ \ \& \ \ |\varphi(\boldsymbol{x})| \leq K \quad (3)$$

φ is the solution of the dual problem (3) and represents the adjoint-source that must be back-propagated within KR FWI (Métivier et al., 2016a). The first constraint on φ is called 1-Lipschitz for the metric c_1. It imposes that changes in φ are sufficiently slow with respect to c_1, which emphasizes low frequencies in φ. This constraint can be reformulated as a local constraint, which leads to a computationally tractable scheme to iteratively solve the discretized eq. (3) even in the multiD case, using the Simultaneous Descent Method of Multipliers (SDMM) method (Métivier et al., 2016b).

We have implemented multiD KR FWI considering the time dimension and a chosen receiver "line" dimension (for instance the inline direction in the marine case). We iteratively re-solve eq. (3) for each line in the chosen receiver-direction with $\boldsymbol{x} = (t_{\boldsymbol{x}}, r_{\boldsymbol{x}}^{line})^{+}$. We pointed out in Poncet et al. (2018) that this approach defines a hybrid misfit mixing to a desired level OT with conventional misfits (namely L_1 and LSQ). We believe this flexibility and the fact that there is a controllable continuum between those misfits is a strong point, see Poncet et al. (2018). In addition, we have greatly optimized our algorithm. The maximum number of SDMM inner iterations to solve the discretized eq. (3) is less than N=30 (like in Métivier et al., 2016b) with a dedicated parallelization.

Parameterization of the multiD metric for an increased continuity in the move-out direction

As in the multiD case, $\boldsymbol{x}$ mixes different physical units, the c_1 metric that appears in the 1-Lipschitz constraint of eq. (3) must be a generalized L_1 distance, parameterized in a so called "Mahalanobis"-like fashion with a covariance in the data space. In our applications, we noticed that accounting for the diagonal part of the covariance is sufficient

$$c_1(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{\sigma_r^{line}} \{ v | t_{\boldsymbol{x}} - t_{\boldsymbol{y}} | + | r_{\boldsymbol{x}}^{line} - r_{\boldsymbol{y}}^{line} | \}$$

where σ_r^{line} represents a variance in the chosen receiver-direction (distance) and $\sigma_t = v/\sigma_r^{line}$ represents a variance in the time direction. v is a velocity that must be parameterized to characterize the average direction along which most correlations between traces occur, i.e. the average move-out direction. A good choice of v is crucial for the success of multiD KR (and certainly of multiD OT in general).

For marine field data (with a mute), Fig. 1 compares a LSQ adjoint-source (or residual) and the corresponding 1D and multiD KR adjoint-sources. The noise in the data is different from one trace to the other, degrading the continuity of LSQ and 1D KR adjoint-sources. This lack of continuity affects the velocity updates. In contrast, multiD KR is able to denoise the adjoint-source, and to increase its coherency in the move-out direction and balance its amplitudes.

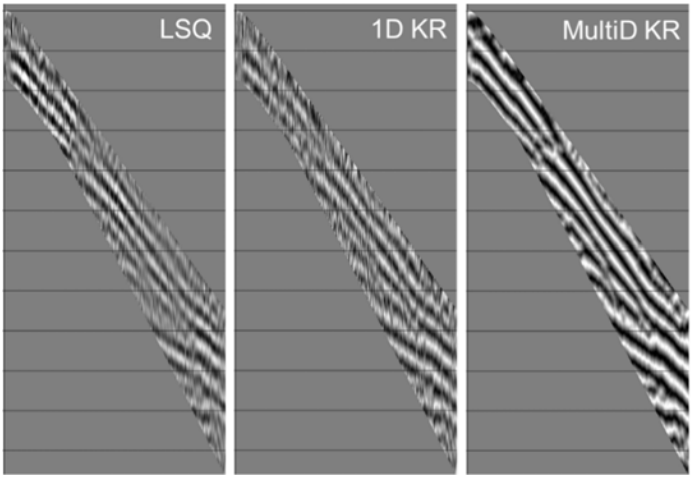

Figure 1: *Real data (3D marine) adjoint-sources at 4 Hz. LSQ FWI (left), 1D KR FWI (middle) and multiD KR FWI (right).*

Application to field datasets

We used $N \leq 30$ inner iterations for each KR problem and a preconditioned L-BFGS optimization scheme for the FWI velocity optimization process.We first present FWI results on a land broadband full azimuth dataset consisting in three separate acquisitions that have been merged, with challenges associated to irregularities for offset distributions among the merged surveys. Fig. 2 compares a LSQ adjoint-source and the corresponding multiD KR adjoint-source. We see that the multiD KR improves the amplitude balancing and coherency in the move-out direction. As careful denoising of the data has been performed

here before FWI, the 1D KR adjoint-source is quite similar to the multiD KR one. But multiD KR nevertheless leads to an improved velocity inversion result as illustrated in Fig. 3, with better structural consistency and well matching. Fig.4 shows that multiD KR FWI gives better alignment to sonic logs at the location of the well and better consistency with the geology compared to LSQ FWI. This can be related to a reduced sensitivity to cycle skipping (Sedova et al., 2018). Note that 1D KR FWI also

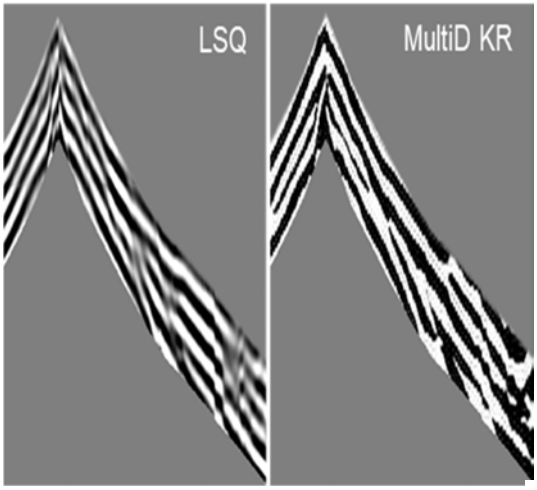

Figure 2: *Real data (3D land) adjoint-sources at 5 Hz. LSQ FWI (left), multiD KR FWI (right)*

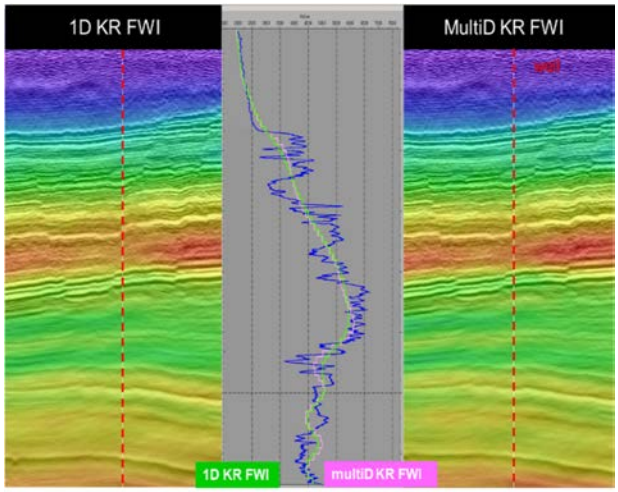

Figure 3: *Benefit of multiD KR FWI (right) compared to 1D KR FWI (left), at 9 Hz.*

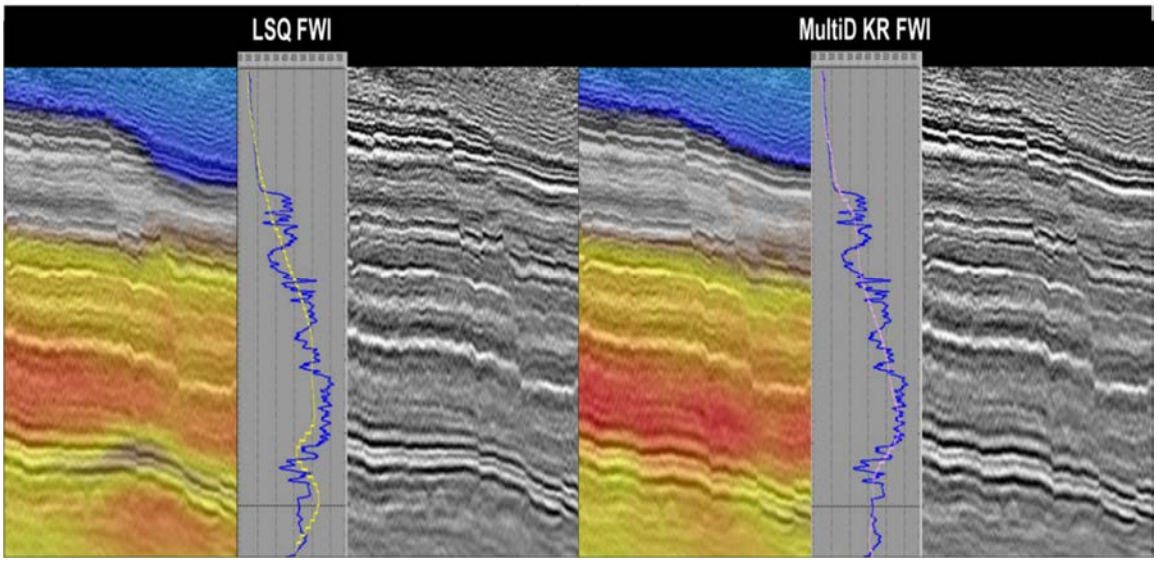

Figure 4: *Results of velocity model building using LSQ FWI (left) and multiD KR FWI (right) at 9 Hz. Three images are presented for each case: the velocity model, the velocity profile and sonic log at the well location, and the migrated stack.*

performs better than LSQ FWI (but less well than MultiD KR FWI as already illustrated).

We finally discuss a marine field data result. Fig. 5 shows that LSQ FWI is cycle-skipped: the indicated "red spots" in the observed data overlaid on top of the modelled data are due to a sudden jump of cycles in the modelled data. This results in a lack of structural consistency and continuity in the inverted velocity. Fig. 5 shows how multiD KR FWI solves for those issues and inverts for an improved velocity.

Conclusions

We demonstrated on field datasets the interest of KR FWI and we discussed the advantages of the multiD (in data space) formulation. We showed that a careful formulation leads to an enhancement of the coherency of the events continuity in the move-out direction and to an improved velocity update

compared to LSQ, that has been illustrated on marine field dataset, but also to 1D KR FWI, that has been illustrated on land field dataset.

For completeness, we mention that Métivier et al. (2018) very recently proposed to use a graph transform of the data within KR FWI to further reduce sensibility to cycle skipping. Because this transform increases the dimensionality of the data space, they proposed to implement it trace-by-trace in KR FWI to keep a reasonable computational cost, leading to a 1D scheme from the original data space point of view. An alternative would be to exploit the denoising power of multiD KR (without data transformation) to start FWI at a lower frequency, also leading to an even more reduced sensibility to cycle skipping.

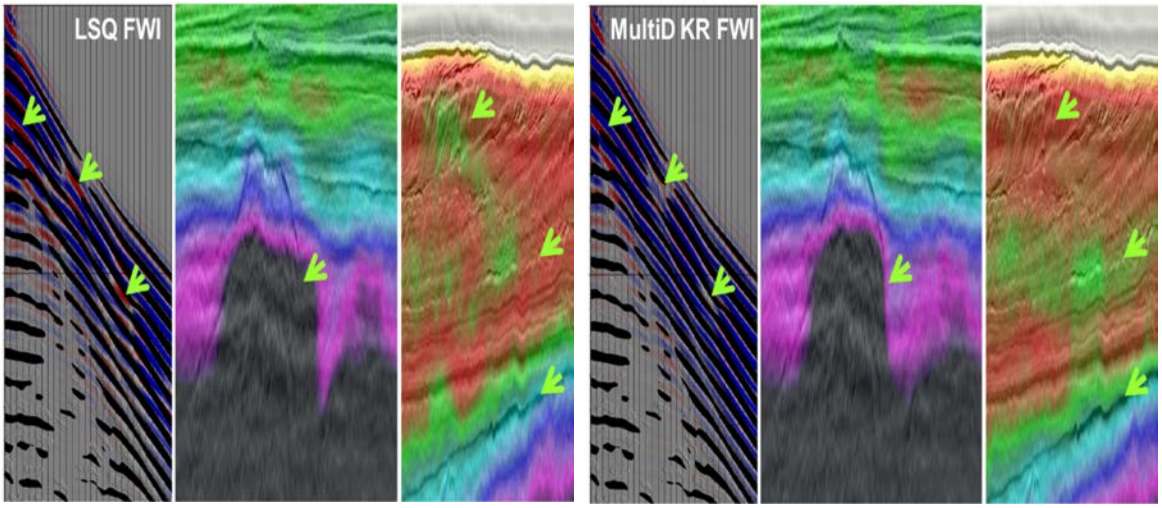

Figure 5: *Results of velocity model building using LSQ FWI (left) and multiD KR FWI (right) at 6 Hz. Three images are presented for each case: observed data (wiggle) overlaid on top of modelled data, the velocity model, and the migrated stack.*

Acknowledgements

We are grateful to Occidental Petroleum, PDO, the Ministry of Oil and Gas of the Sultanate of Oman, INEOS and CGG their permission to publish this work. We are grateful to Hervé Prigent, Ghislain Viguier, Sylvain Masclet, Milad Bader, Krzysztof Ubik and Gilles Lambaré for their involvement.

References

Engquist, B. and Froese, B. D. [2014] Application of the Wasserstein metric to seismic signals. *Communications in Mathematical Sciences* **12**, 979 – 988.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E. and Virieux, J. [2016a] Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophysical Journal International* **205**, 345-377.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E. and Virieux, J. [2016b] An optimal transport approach for seismic tomography: application to 3D FWI. *Inverse Problems* **32**, 115008.

Métivier, L., Alain, A., Brossier, R., Mérigot, Q., Oudet, E. and Virieux, J. [2018] A graph-space approach to optimal transport for full waveform inversion. *87th SEG Expanded Abstracts*, p. 1158.

Poncet, R., Messud, J., Bader, M., Lambaré, G., Viguier, G. and Hidalgo, C. [2018] FWI with optimal transport: a 3D implementaion and an application on a field dataset. *80th EAGE Conference and Exhibition*, EAGE, Expanded Abstract.

Qiu, L., Ramos-Martínez, J., Valenciano, A., Yang, Y. and Engquist, B. [2017] FWI with an exponentially-encoded optimal transport norm. *87th SEG Expanded Abstracts*, p. 1286.

Sedova, A., Messud, J., Prigent, H., Masclet, S., Royle, G. and Lambaré, G. [2018] Acoustic FWI: an efficient model-building tool for broadband land data. *SEG/KOC workshop in Kuwait*, Expanded Abstract.

Yang, Y., Engquist, B., Sun, J. and Froese, B. D. [2018] Application of Optimal Transport and the Quadratic Wasserstein Metric to Full-Waveform Inversion. *Geophysics* **83**, R43-R62.