2021 EAGE ANNUAL
82ND CONFERENCE & EXHIBITION
AMSTERDAM | THE NETHERLANDS

OPTIMAL TRANSPORT FWI WITH GRAPH TRANSFORM: ANALYSIS AND PROPOSAL OF A PARTIAL SHIFT STRATEGY

F. Kpadonou[1], J. Messud[1], A. Sedova[1], M. Reinier[1]

[1] CGG

Summary

The use of graph-space optimal transport within full-waveform inversion (FWI) has recently been proposed to enhance robustness to cycle-skipping. In this paper, we discuss several features of the graph-space optimal transport FWI, emphasizing in particular its ability to detect time shifts between modelled and observed data and its characteristics in terms of adjoint source (the signal back-propagated from the receiver side during the iterative optimization process). Our analysis is driven by finding an optimal adjoint-source that is robust to cycle-skipping. This leads us to propose an alternative graph-space-inspired scheme called enhanced kinematic transform FWI. Our approach involves a modification of the adjoint-source introducing a partial shift strategy, allowing to highlight the kinematic information. We illustrate the enhanced robustness with respect to cycle-skipping on synthetic and field datasets with comparison to conventional FWI, Kantorovich-Rubinstein optimal transport FWI and graph-space optimal transport FWI.

Optimal transport FWI with graph transform: Analysis and proposal of a partial shift strategy

Introduction

Conventional full-waveform inversion (FWI) is based on a least squares (LS) misfit function, which has proven to be effective for high-resolution velocity model building (VMB). However, LS FWI is plagued by cycle-skipping, so that the local optimization process requires starting from a good initial model. Many alternative misfit functions have been proposed to mitigate this issue. Among those, misfits based on the optimal transport (OT) theory recently aroused attention in geophysics. Métivier et al. (2016), proposed to use the Kantorovich-Rubinstein (KR) norm-based version of OT, inspired from a 1-Wasserstein distance, which allows considering seismic data directly without any transformation. Encouraging results were obtained on several real datasets using KR in FWI (see for example, Messud et al. (2019), Sedova et al. (2019) and Hermant et al. (2020)).

More recently, Métivier et al. (2018) proposed a complementary OT version based on a 2-Wasserstein distance together with a specific positive transformation of the data called "graph transform". The resulting graph space (GS) OT formalism allows to somewhat "shift" events contained in the data (whereas KR enhances the kinematics information without shifting the events), in the spirit of dynamic warping FWI (Wang et al., 2016), giving robustness with respect to cycle-skipping. Encouraging results were obtained by Métivier et al. (2018), Provenzano et al. (2020) and Pladys et al. (2020).

Following the latter work, we discuss in this article several features of GS FWI. We emphasize how GS OT allows to detect a change of kinematics between the modelled and observed data, and that its robustness to cycle-skipping seems equivalent to that of KR FWI. We then propose a partial shift strategy and the "enhanced kinematic transform" (EKT) method to further focus the adjoint-source on the kinematic information, and obtain yet more robustness to cycle-skipping. We illustrate our points on synthetic and field datasets.

Classical LS and OT (KR, GS) misfits

The data space is discretized through a vector $\boldsymbol{x}:=\{x_i, i=1\cdots M\}$ that describes the time and receiver positions. The dependency to shot position is assumed implicitly. For each shot, we denote $\boldsymbol{d}_{obs}:=\{d_{obs}(x_i), i=1\cdots M\}$ the observed data vector and $\boldsymbol{d}:=\{d(x_i), i=1\cdots M\}$ the data vector modelled using a subsurface model. For a misfit $J(\boldsymbol{d}_{obs},\boldsymbol{d})$, the data-space gradient $\delta J/\delta \boldsymbol{d}$ defines the adjoint-source, translated into the velocity-space gradient via the adjoint-state method. The LS misfit is

$$J_{L_2}(\boldsymbol{d}_{obs},\boldsymbol{d}) = \frac{1}{2}\Delta\boldsymbol{d}\cdot\Delta\boldsymbol{d} \qquad \text{where} \quad \Delta\boldsymbol{d}=\boldsymbol{d}_{obs}-\boldsymbol{d}. \tag{1}$$

The LS adjoint-source equals the data residual $\Delta\boldsymbol{d}$. The KR adjoint-source can be viewed as a smart processing of $\Delta\boldsymbol{d}$ that gives robustness to cycle-skipping by enhancing the low-frequency content and reducing the amplitude dynamics. It is defined as the *argmax* of the problem (Métivier et al, 2016b)

$$J_{KR}(\boldsymbol{d}_{obs},\boldsymbol{d}) = \max_{\boldsymbol{\varphi}} \boldsymbol{\varphi}.\Delta\boldsymbol{d} \;\; \text{s.t.} \;\; \left|\varphi(x_i)-\varphi(x_j)\right| \le \left\|x_i-x_j\right\|_1 \;\; \& \;\; \left|\varphi(x_i)\right| \le K. \tag{2}$$

Complementarily, Métivier et al. (2018) proposed the GS OT formulation based on the 2-Wasserstein distance together with a specific positive transformation of the data called "graph transform", where data amplitudes are nested with the corresponding data positions. For computational purposes, each data trace is considered independently; hence we consider now the discrete time positions denoted by $\boldsymbol{t}:=\{t_i, i=1\ldots N\}$ instead of $\boldsymbol{x}$. The modelled and observed data traces are viewed as two clouds of points $(t_i, d(t_i))$ and $\left(t_j, d_{obs}(t_j)\right)$ of uniform mass. The OT problem can then be recast into

$$J_{GS}\left((\boldsymbol{t},\boldsymbol{d}_{obs}),(\boldsymbol{t},\boldsymbol{d})\right) = \min_{\boldsymbol{\sigma}\in S(N)} \sum_{i=1,N} c_{i\sigma_i}, \quad \text{with} \quad 2\,c_{ij} = \left|t_j-t_i\right|^2/t_N^2 + \alpha^2\left|d_{obs}(t_j)-d(t_i)\right|^2. \tag{3}$$

$S(N)$ denotes the set of permutation of a vector of dimension N and c_{ij} the transportation cost from the modelled data sample i to the observed data sample j. The weighting parameter α is used to control the amount of data displacement along the time axis. Denoting by $\boldsymbol{\sigma}^*$ the optimal permutation, computed using the auction algorithm of Bertsekas et al. (1989), the GS adjoint-source reads (Métivier et al. 2018)

$$\frac{\delta J_{GS}}{\delta \boldsymbol{d}} = \alpha^2\,\Delta\tilde{\boldsymbol{d}} \qquad \text{where} \qquad \Delta\tilde{\boldsymbol{d}}=\tilde{\boldsymbol{d}}_{kin}-\boldsymbol{d} \qquad \text{and} \quad \tilde{d}_{kin}(t_i):=d_{obs}(t_{\sigma_i^*}). \tag{4}$$

Interestingly, the GS adjoint-source residual $\Delta\tilde{\boldsymbol{d}}$ looks like the LS residual $\Delta\boldsymbol{d}$, with the difference that a time or kinematic transformation of the observed data has been performed. In other terms, events

contained in the data are "shifted", which gives robustness to cycle-skipping, in the spirit of dynamic warping FWI (Wang et al., 2016).

Conventional GS FWI interest and problematics

The main GS parameter α plays a crucial role. For $\alpha \to 0$ the minimization problem in Equation 3 leads to an adjoint-source residual that tends to the LS residual, i.e. $\Delta\tilde{\boldsymbol{d}} \to \Delta\boldsymbol{d}$. The other extreme case, $\alpha \to \infty$, allows a large displacement of data along the time axis in order to reduce the amplitude term ($|d_{obs} - d|^2$) in Equation 3 (Métivier et al., 2018) but the adjoint-source residual $\Delta\tilde{\boldsymbol{d}}$ in Equation 4 then tends to vanish and become useless (see Figure 1c). To avoid this pathological behaviour, we observed that the parameter α must be chosen very small in practice, as concluded in Métivier et al. (2018), leading to a "perturbation" of the LS residual shown in Figures 1a and 1b.

We tested the GS scheme on the 2D Marmousi 2 dataset created in the constant-density acoustic approximation with a Ricker source centred on 6 Hz. We used a preconditioned L-BFGS optimization scheme for the FWI optimization process. We run 30 FWI iterations performed directly at 10 Hz, keeping all the data (first break, reflections, multiples…). Starting with an initial velocity model where LS is cycle-skipped, Figure 2a shows that both KR and GS overcome this issue. However, in the more difficult configuration of Figure 3 where a perturbation is added to the shallow layer in the initial model, both GS and KR fail. Interestingly, we observed in our tests a comparable robustness to cycle-skipping between KR and GS.

Beyond conventional GS using a partial shift, and EKT scheme

We now intend to address the pathological situation of vanishing GS adjoint-source residual $\Delta\tilde{\boldsymbol{d}}$ for large amount of data displacement along the time axis value, i.e. large value of α. The hope is to become even more robust to cycle-skipping since the valley of convexity of the GS misfit with respect to time shift enlarges with α (illustration in Métivier et al. (2018) for two shifted Ricker wavelets). The strategy that we propose is to change the definition of $\tilde{d}_{kin}$ in the adjoint-source residual $\Delta\tilde{\boldsymbol{d}}$ of Equation 4 by

$$\tilde{d}_{kin}\left(t_i + \beta\left(t_{\sigma_i^*} - t_i\right)\right) = d_{obs}\left(t_{\sigma_i^*}\right). \quad (5)$$

Taking $\beta = 0$ would lead to the conventional GS adjoint-source residual. Taking $\beta = 1$ would lead to the LS residual. Interestingly, a choice of $\beta \in (0,1)$ leads to a "partial shift" and makes clear that GS defines a form of kinematic transform (KT), in the spirit of dynamic warping FWI (Wang et al., 2016). Considering a large value for α is now possible without creating a null $\Delta\tilde{\boldsymbol{d}}$ thanks to the partial shift ratio β. This is illustrated in Figure 1d in contrast to Figure 1c.

To enhance even more the KT, we propose another definition of $\tilde{d}_{kin}$ in Equations 4 and 5 based on the synthetic data (still considering $\Delta\tilde{\boldsymbol{d}}$ in Equation 4 as the adjoint-source residual to be back-propagated)

$$\tilde{d}_{kin}\left(t_{\sigma_i^*} + \beta\left(t_i - t_{\sigma_i^*}\right)\right) = d(t_i). \quad (6)$$

For $\beta = 0$, this definition involves applying the transpose of the permutation vector to the modelled data (and not the permutation vector to the observed data like in Equation 5; indeed σ^* only occurs in the left hand side of Equation 6). A choice of $\beta \in (0,1)$ allows a partial shift strategy. The underlying idea is to use an adjoint-source residual $\Delta\tilde{\boldsymbol{d}}$ that focuses more on the kinematics change comparing shifts between similar data (here the modelled one) and not on events missing in the modelled data, as illustrated in Figure 1e. The aim is to be more robust to cycle-skipping, especially at the very first FWI iterations. We call this GS-derived approach the EKT scheme and will give formal grounds to it (and to the point in next paragraph) in a further article.

Note that the GS or EKT approaches can naturally be embedded in KR, replacing the residual $\Delta\boldsymbol{d}$ in Equation 2 by the GS "residual" $\Delta\tilde{\boldsymbol{d}}$ of Equation 4. One interesting advantage, by doing so, is to add the features of KR to EKT: enhanced low frequency content, amplitude balancing, stronger continuity in the move-out direction (KR) and "shifts" of events contained in the data (EKT), see Figure 1e, for even more robustness to cycle-skipping. Figure 2b shows that the EKT approach (embedded in KR) overcomes cycle-skipping better than GS and KR only on Marmousi when starting the FWI with the more difficult initial model.

Application to a field dataset

The Greater Castberg area of the Barents Sea is challenging for VMB: it contains faulted structures in the deeper part of the section and a lot of gas hydrates and chimneys in the shallow part. We used data from an acquisition performed with 8 km long variable depth streamers, delivering penetration of diving

waves down to 2.5 km. The FWI was run from 3 to 7 Hz. We started the update from an initial velocity model including gas pocket anomalies in the shallow (first 600 m). The deeper part however remains challenging, with an initial model that is smooth and inaccurate. Figure 3a compares GS FWI with KR and LS FWI, starting from the initial model (a). LS FWI fails to update the model correctly. Both KR and GS give a similar realistic velocity result providing a good fit between observed and modelled data. In a second test, we changed the initial model (a) by increasing slightly velocities in the deeper part (by adding a perturbation – see Figure 3b), making it harder for FWI to avoid the cycle-skipping. Figure 3b shows that LS, GS (and also KR not shown here) fail. On the other hand, the use of EKT at 3 Hz followed by GS from 4 to 7 Hz has provided a stable and a geologically-consistent resulting model (we observed in our tests that using EKT at the very first FWI iterations was sufficient for cycle-skipping).

Conclusion

We presented a GS FWI implementation and noticed on synthetic and field data tests that the robustness to cycle-skipping seems equivalent to KR FWI. To go beyond this, we introduced a partial shift strategy within GS, leading to the kinematic transform concept. On this basis, we proposed the EKT method to better focus the adjoint-source on the kinematic information. We have illustrated on synthetic and field datasets the interest to use EKT at the very first FWI iterations for more robustness to cycle-skipping.

Acknowledgements

We are grateful to CGG, CGG Multi-Client and TGS for their permission to publish this work, and to Gilles Lambaré, Paolo Poggi, Loïc Janot and Hervé Prigent for enlightening discussions.

References

Bertsekas, D. P. and Castanon, D. [1989] The auction algorithm for the transportation problem. *Annals of Operations Research*, **20**, 67–96.

Hermant, O., Aziz, A., Warzocha, S. and Al Jahdhami, M. [2020] Imaging Complex Fault Structures On-shore Oman Using Optimal Transport Full Waveform Inversion. *EAGE 2020 Annual Conference and Exhibition, Extended abstract,* p. 1-5.

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

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.

Messud, J. and Sedova, A. [2019] Multidimensional Optimal Transport for 3D FWI: Demonstration on Field Data. *81st EAGE Conference and Exhibition*, Extended abstract, Tu R08 02.

Pladys, A., Brossier, R., Irnaka, M. Kamath, N. and Métivier, L. [2020], Graph space optimal transport based FWI: 3D OBC Valhall case study, *SEG Technical Program Expanded Abstracts*, 696-700.

Provenzano, G., Brossier, R., Métivier, L. and Li, Y. [2020], Joint FWI of diving and reflected waves using a graph space optimal transport distance: Synthetic tests on limited-offset surface seismic data, *SEG Technical Program Expanded Abstracts*, 780-784

Sedova, A., Messud, J., Prigent, H., Masclet, S., Royle, G. and Lambaré, G. [2019] Acoustic Land Full Waveform Inversion on a Broadband Land Dataset: the Impact of Optimal Transport. *81st EAGE Conference and Exhibition, Extended abstract*, Tu R08 07.

Wang, M., Xie, Y., Xu, W.Q., Xin, K. F., Chuah, B. L., Loh, F. C., Manning, T., Wolfarth, S. [2016] Dynamic-warping FWI to overcome cycle skipping. *86th SEG Expanded Abstracts,* 1273.

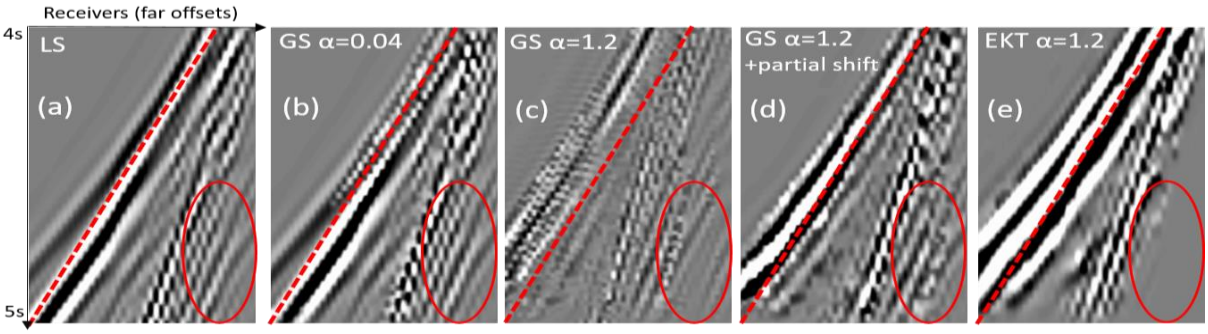

Figure 1 *Different Marmousi adjoint-sources, zoomed in the diving waves area, in the initial model of Figure 2 (b). GS (α =0.04) only slightly changes the kinematics of the events compared to the LS residual. The partial shift approaches (α =1.2, β=0.5) produce more changes in the kinematics of the events, and EKT allows to put "more weight" on the diving waves.*

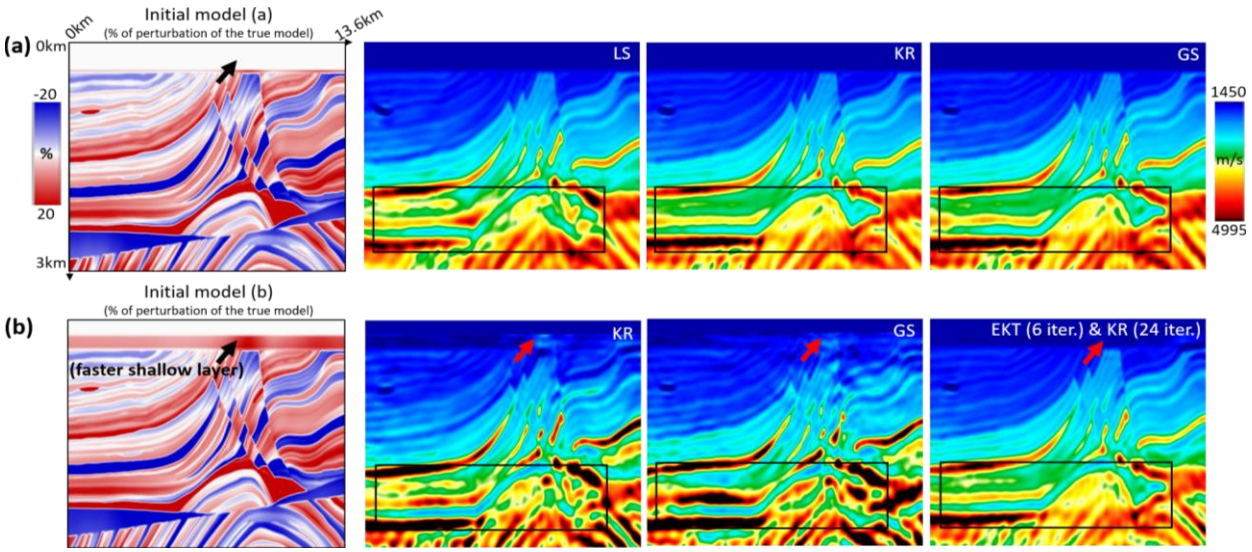

Figure 2 *Marmousi 2 velocity models. 30 FWI iterations performed directly at 10Hz, keeping all the data. (a) Starting the FWI with an initial model where LS is cycle-skipped. (b) Starting the FWI with an initial model more perturbed in the shallow layer (underlined by the arrows) where LS, KR and GS are cycle-skipped. Initial models are represented in percentage of perturbation of the true model.*

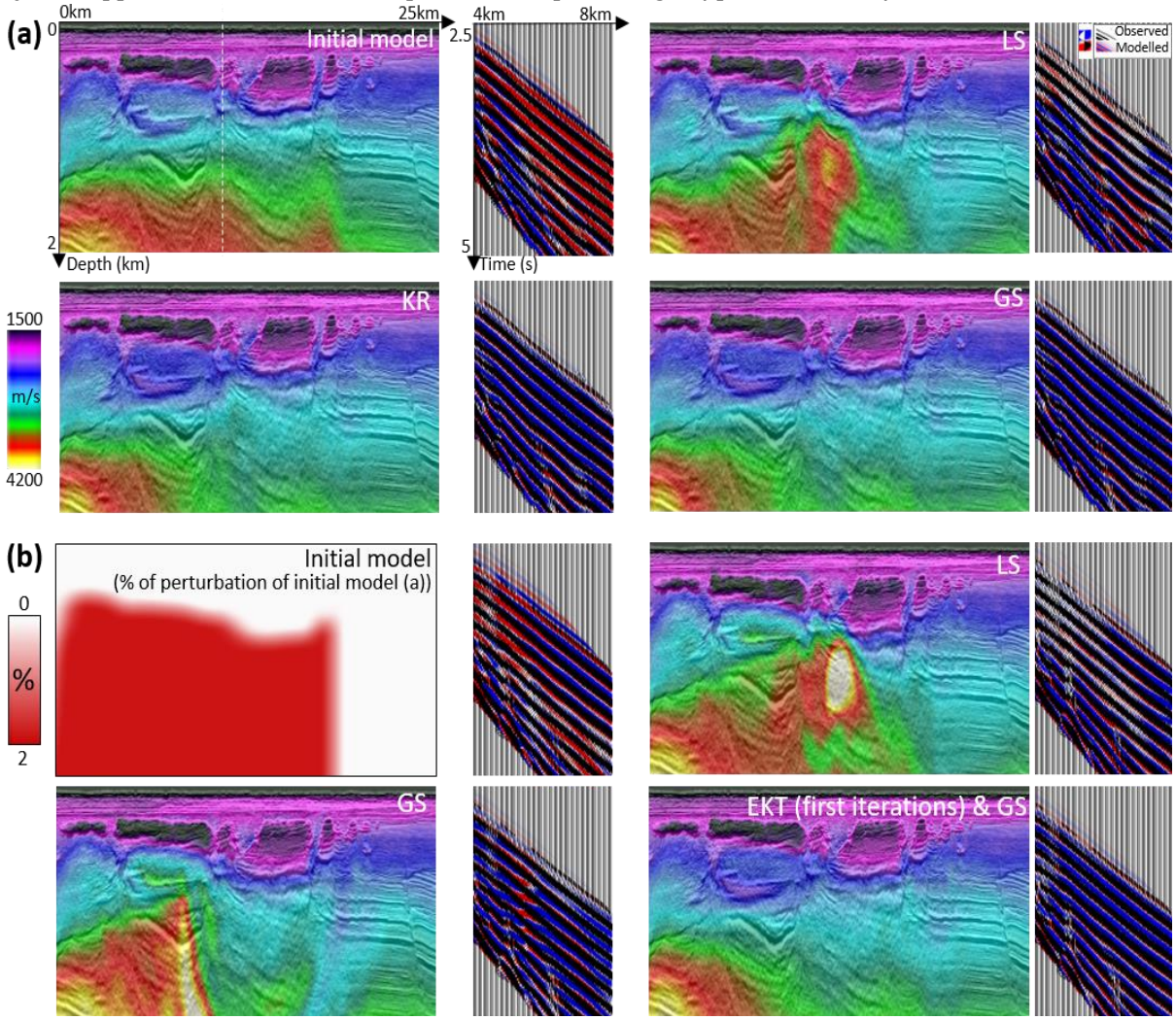

Figure 3 *Inverted velocity models by 3 to 7 Hz FWI and data fit QCs (with shot position at the white line). (a) Starting the FWI with a model which contains already gas pocket anomalies in the shallow. LS is cycle-skipped, whereas GS and KR are not. (b) Starting the FWI with a model where the velocity has been slightly increased in the deeper part compared to initial model (a), LS, GS (and KR not shown here) are cycle-skipped, whereas starting with EKT is not.*