Tu_R08_03

Graph Space Optimal Transport for FWI: Auction Algorithm, Application to the 2D Valhall Case Study

L. Metivier[1,2]*, R. Brossier[2], Q. Merigot[3], E. Oudet[1]

[1] CNRS, Univ Grenoble Alpes, LJK; [2] Univ. Grenoble Alpes, ISTerre; [3] Univ. Paris Sud, LMO

Summary

Optimal transport based misfit functions have been introduced to mitigate cycle skipping in full waveform inversion. Recently, we have proposed to compare the discrete graph of the data through optimal transport. This strategy makes possible to interpret non-positive data through optimal transport (OT), while maintaining its convexity property with respect to time and amplitude shifts. Here, we introduce a novel numerical strategy, which makes the computational overcost associated with the graph space OT approach almost negligible compared to a conventional least-squares distance. This is possible through the interpretation of the graph space OT problem as a linear sum assignment problem, for which specific algorithms exist: we select the auction algorithm for its good performance on small scale, dense instances of such problems. We introduce a tuning parameter, based on the estimation of the maximum time shift, to scale the misfit function properly. Application on visco-acoustic synthetic data illustrate the interest of the strategy. From significantly cycle-skipped data, the method retrieves a meaningful estimation of the velocity model, while the conventional least-squares approach and one previously introduced OT based approach both fail. The computational cost increase is only few percents compared to the least-squares misfit function.

Introduction

Despite many successful applications of full waveform inversion (FWI) for velocity model building at the exploration scale, this iterative data matching method still suffers from limitations related to the non-convexity of the misfit function which is minimized through local optimization. This non-convexity is mostly due to the least-squares distance, which is conventionally used to measure the discrepancy between observed and calculated data. This distance is non-convex with respect to the time-shifts produced by the subsurface velocity perturbations that FWI aims at reconstructing (Virieux and Operto, 2009).

Many researches have focus on modifying the conventional FWI to overcome this difficulty, and amongst them, the use of optimal transport (OT) distance. One appealing feature is that the OT distance is convex with respect to shifted patterns in the compared quantities. However, it applies to positive and normalized signals. Applications to oscillatory seismic data require specific strategies, either relying on non-linear transforms and normalizations, or on a specific instance of OT, named Kantorovich-Rubinstein (KR) distance, which can be extended to the comparison of non-positive data (Métivier et al., 2016). These different strategies proposed in the literature have been recently reviewed in Métivier et al. (2018), and it appears they are either difficult to be applied to real data, either lose the convexity property of OT. This conclusion has motivated Métivier et al. (2018) to propose a graph space OT strategy, in which the discrete graph of the data is compared through OT. This yields a feasible OT problem regarding positivity, normalization, and also preserves the convexity property. However, the computational cost drastically increases, as a two-dimensional OT problem has to be solved for each seismic trace.

In the present study, we propose a novel numerical strategy for the graph space OT approach, which makes possible to consider large and realistic size FWI problems. The approach relies on the formulation of the OT distance between discrete graph as a linear assignment sum problem (LSAP). The auction algorithm is used to solve it: for small scale, dense instance of LSAP, this algorithm is very efficient (Bertsekas and Castanon, 1989). To demonstrate its efficiency, the method is applied to a realistic 2D visco-acoustic FWI case. It is compared with a standard L^2 distance and the OT-KR distance. The graph space OT outperforms both approaches. Thanks to the efficiency of the auction algorithm, the computational increase with respect to the L^2 distance is only few percent.

Discrete graph space OT as a LSAP problem

Consider a seismic trace $d(t)$ discretized as $(d_1,\ldots,d_N)$. We denote its discrete graphs by $(t,d) = ((t_1,d_1),\ldots,(t_N,d_N)) \in \mathscr{R}^{2N}$. Let d_{cal}, d_{obs} be a calculated and an observed trace. The 2-Wasserstein distance between (t,d_{cal}) and (t,d_{obs}), considered as ensembles of N delta Dirac points in a 2D space, writes

$$\begin{aligned} W_2^2\left((t,d_{cal}),(t,d_{obs})\right) = \min_{\gamma} \quad & \sum_{ij=1}^{N} \gamma_{ij}\left(|t_i-t_j|^2 + |d_{cal,i}-d_{obs,j}|^2\right), \\ & \gamma_{ij} \geq 0, \; \sum_{i=1}^{N}\gamma_{ij} = 1, \; \sum_{j=1}^{N}\gamma_{ij} = 1, \end{aligned} \tag{1}$$

(see for instance Benamou and Brenier, 2000, for more details). The problem (1) is a Linear Sum Assignment Problem (LSAP), a specific linear programming problem which has been intensively studied in the second half of the 20th century for its use in the modeling of many economy problems. A solution $\overline{\gamma}$ to (1) always exists and corresponds to a permutation matrix: a single coefficient per row is equal to 1, the others are equal to 0 (Birkhoff, 1946). The solution of (1) amounts to finding a permutation matrix which minimizes the transport cost of (t,d_{cal}) to (t,d_{obs}) where each displacement is measured with a L^2 distance in the time/amplitude 2D space.

Denoting $\overline{\sigma}$ the permutation associated with $\overline{\gamma}$, we introduce the graph space OT misfit function as

$$h_2(d_{cal},d_{obs}) = \sum_{i=1}^{N} c_{i,\overline{\sigma}(i)}, \tag{2}$$

where $c_{ij} = |t_i-t_j|^2 + |d_{cal,i}-d_{obs,j}|^2$. For practical applications, it is necessary to scale properly the ratio between the time and amplitude axis. Denote T the maximum time, and A the peak amplitude difference

$$A = \max\left(\max_{ij}|d_{cal,i}-d_{cal,j}|, \max_{ij}|d_{obs,i}-d_{obs,j}|\right). \tag{3}$$

If $T >> A$, it is cheaper to transport the points along the amplitude axis. In this case, the solution $\overline{\sigma}$ is the identity, and (2) amounts to the L^2 distance. If $T << A$, it is cheaper to displace the points along the

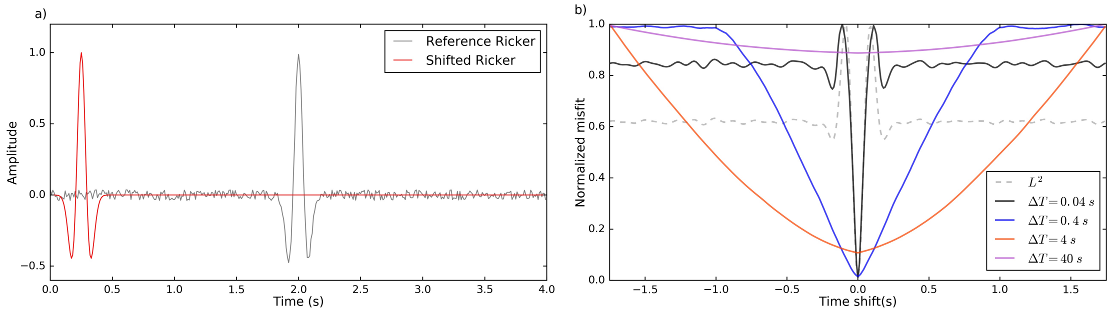

Figure 1: a) reference Ricker function (solid gray line), shifted in time Ricker function (solid black line). b) h_2 misfit depending on the time shift, for different values of ΔT. The reference L^2 misfit is in dashed gray line. Misfit function for $\Delta T = 0.04$ s (solid black line), $\Delta T = 0.4$ s (solid blue line), $\Delta T = 4$ s (solid orange line), $\Delta T = 40$ s (solid purple line).

time axis. If the amplitude of the signal can be predicted with infinite accuracy, the measure (2) depends exclusively on the time differences $|t_i - t_{\overline{\sigma}(i)}|^2$. However, in practice, the amplitude is never perfectly predicted, and the measure (2) would be dominated in this case by the amplitude mismatch.

For this reason, we introduce a control scaling factor ΔT such that

$$c_{ij} = |t_i - t_j|^2 + |\frac{\Delta T}{A}\left(d_{cal,i} - d_{obs,j}\right)|^2. \tag{4}$$

The control ΔT can be interpreted as the maximum time shift one wants to recover from OT. For instance if $\Delta T = T$, there is a perfect balance between time and amplitude axis, and the cost of displacing one point along the whole time axis is the same as displacing it along the whole amplitude axis. In practice, ΔT is taken as the largest time-shift we can expect from the initial model, estimated through seismic modeling and data analysis.

We measure the h_2 distance between a reference and a shifted in time Ricker signal, depending on the time shift. Noise is added to the reference Ricker so that the amplitude cannot be predicted accurately (Fig. 1a). Increasing values of ΔT makes h_2 evolve from a quasi L^2 misfit function to a convex misfit function with respect to the time shift. However, choosing too large values for ΔT ends up in an almost constant misfit function: the amplitude mismatch due to the noise starts dominating the misfit measurement (Fig. 1b).

For practical FWI problems, with N_s shot gathers containing N_r traces denoted by $d_{obs}^{s,r}$ and $d_{cal}^{s,r}$, we consider the following graph space OT FWI problem

$$\min_m f[m] = \sum_{s=1}^{N_s}\sum_{r=1}^{N_r} w^{s,r}\, h_2\left(d_{cal}^{s,r}[m], d_{obs}^{s,r}\right). \tag{5}$$

We introduce the weights $w^{s,r} = ||d_{obs}^{s,r}||^2$ to restore the AVO information in our misfit function. Here, $d_{cal}^{s,r}[m]$ denotes the synthetic trace for source s and receiver r computed through the solution of partial differential equations (PDE) describing the seismic wave propagation, parameterized by m. The FWI problem consists in estimating m.

Solving (5) requires to access the gradient of $f[m]$. Through the adjoint state technique (Plessix, 2006), this requires to compute an adjoint source μ. We can show that in non singular case the adjoint source depends only on $\overline{\sigma}$, such that

$$\mu_i^{s,r} = \frac{\partial h_2}{\partial d_{cal,i}^{s,r}} = 2\left(d_{cal,i}^{s,r} - d_{obs,\overline{\sigma}(i)}^{s,r}\right),\ \ i = 1,\ldots,N. \tag{6}$$

This displays a close connection with the L^2 distance: the adjoint source is equal to the residual between calculated and observed data, at the notable exception that the comparison is not made sample by sample, but between couple of samples $(i, \overline{\sigma}(i))$ given by the optimal assignment from the graph space OT comparison.

For the numerical solution of (1), we rely on the auction algorithm, in its Gauss-Seidel version (Bertsekas and Castanon, 1989). This algorithm has a cubic complexity ($O(N^3)$). However, for small instances ($N = O(10^2)$ to $N = O(10^3)$), the method exhibits fast computational time (from 8×10^{-3} s to 0.38 s on a standard laptop). In the frame of realistic FWI applications, relying on a Nyquist sampling of the seismic traces yields a number of discretization points in time within this range. This makes possible to use the misfit function $f[m]$ for realistic size FWI problems.

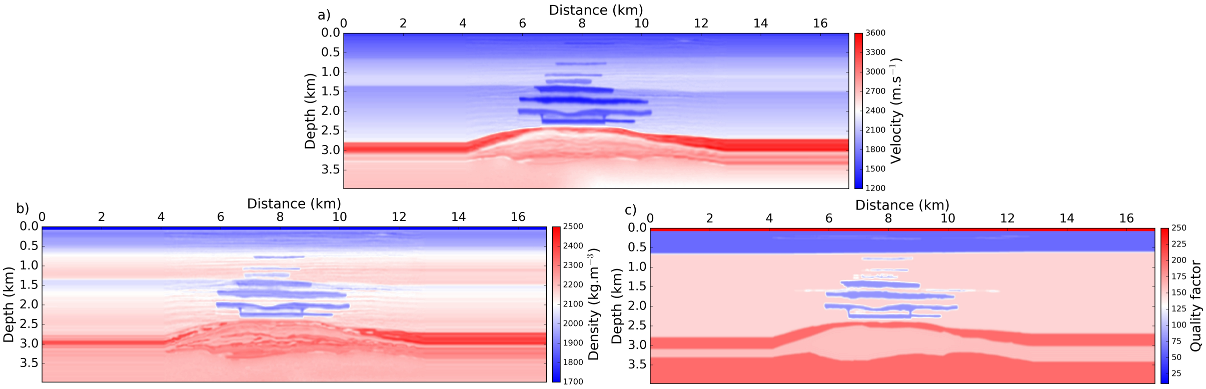

Figure 2: 2D Valhall model for a) velocity $V_P(x,z)$, b) density $\rho(x,z)$, and c) quality factor $Q_P(x,z)$.

Application to the 2D Valhall visco-acoustic case study

We apply the strategy described above on the 2D Valhall synthetic model in Figure 2. We construct a set of observations with a visco-acoustic variable density modeling engine (Yang et al., 2018). We use a fixed-spread surface acquisition with 128 sources and 169 receivers deployed at 50 m depth, with an interval of 130 m between sources and 100 m between receivers. The time signature of the source is a low-cut filtered Ricker with no energy below 2.5 Hz. The total recording time is 8 s. A Gaussian white noise is added, such that the SNR of the data is equal to 10. The inversion is performed with a depth preconditioned l-BFGS approach, with maximum 500 iterations allowed. A Gaussian smoothing of the gradient is used, with correlation lengths depending on the estimated local resolution.

The initial velocity model $V_{P,0}$ is built through a Gaussian smoothing of the exact model, with correlation lengths of 1.25 km in x and z directions. The maximum time shift between observed and calculated data is estimated to $\Delta T = 0.4$ s in $V_{P,0}$: we use this value for the graph space OT approach. The initial density model ρ_0 is built from a Gardner law. The initial quality factor $Q_{P,0}$ is equal to 1000 in the water layer, and 100 below. A mono-parameter V_P inversion is performed, with ρ_0 and $Q_{P,0}$ as passive parameters.

We compare the results obtained using a L^2 distance, the OT-KR approach, and the proposed graph space OT method in Figure 3. The L^2 and OT-KR inversions fail to recover a meaningful estimation of the velocity due to strong cycle skipping in $V_{P,0}$. The graph space OT approach is able to reconstruct a more satisfactory model. We stress here that the amplitude of the signal cannot be predicted accurately due to noise and missing information on density and attenuation models: we are not in an inverse crime setting. The data computed in the initial and final models show that only using the graph space OT approach, FWI is able to reconstruct correctly the intermediate to large offsets, associated with waves traveling through the gas layers. These are strongly cycle skipped in the initial model.

The proposed approach induces an overcost of approximately 6.5% of the total computational cost on the parallel cluster we have used for applications (note that it is 13% for the OT-KR approach). This makes the method highly efficient compared to the strategy proposed in Métivier et al. (2018), and therefore applicable for large and realistic 3D targets.

Conclusion and perspectives

The graph space OT approach based on the auction algorithm seems interesting to mitigate cycle skipping. It allows to compare non-positive data through OT while maintaining the convexity with respect to time shifts, inducing an almost negligible computational cost increase compared to a L^2 misfit. An easy-to-tune scaling approach is provided here to adapt the misfit function to the time shifts that need to be accounted for, depending on the quality of the initial model. Satisfactory results obtained on a realistic 2D Valhall visco-acoustic model prompt us to investigate the use of this strategy on field data, which will be the matter of future studies. A more detailed analysis of this method and its application will be submitted soon (Métivier et al., 2019).

Acknowledgements

This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC, and TOTAL. This study was granted access to the HPC resources of CIMENT infrastructure (https://ciment.ujf-grenoble.fr) and CINES/IDRIS/TGCC under the allocation 046091 made by GENCI.

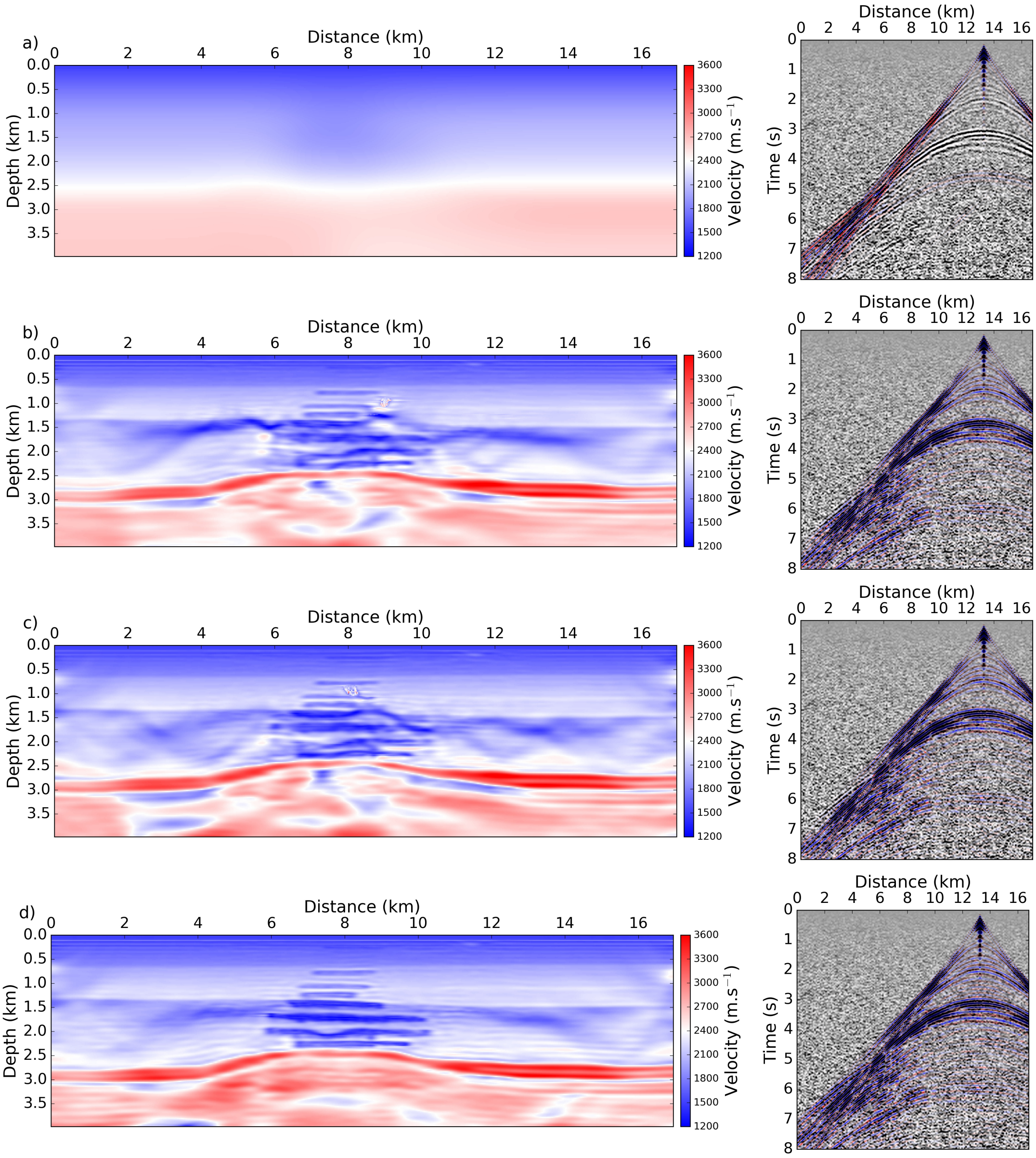

Figure 3: a) Initial V_p model, comparison between exact (black and white) and modeled data (blue and red) . b) L^2 result, corresponding data comparison. c) KR result, corresponding data comparison. d) Graph space OT result, corresponding data comparison.

References

Benamou, J.D. and Brenier, Y. [2000] A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, **84**(3), 375–393.

Bertsekas, D.P. and Castanon, D. [1989] The auction algorithm for the transportation problem. *Annals of Operations Research*, **20**(1), 67–96.

Birkhoff, G. [1946] Three observations on linear algebra. *Univ. Nac. Tucumán*, **A**(5), 147–151.

Métivier, L., Allain, A., Brossier, R., Mérigot, Q., Oudet, E. and Virieux, J. [2018] Optimal transport for mitigating cycle skipping in full waveform inversion: a graph space transform approach. *Geophysics*, **83**(5), R515–R540.

Métivier, L., Brossier, R., Mérigot, Q. and Oudet, E. [2019] A graph space optimal transport distance as a generalization of L^p distances: application to a seismic imaging inverse problem. *Inverse Problems*, **to be submitted**.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E. and Virieux, J. [2016] An optimal transport approach for seismic tomography: Application to 3D full waveform inversion. *Inverse Problems*, **32**(11), 115008.

Plessix, R.E. [2006] A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, **167**(2), 495–503.

Virieux, J. and Operto, S. [2009] An overview of full waveform inversion in exploration geophysics. *Geophysics*, **74**(6), WCC1–WCC26.

Yang, P., Brossier, R., Métivier, L., Virieux, J. and Zhou, W. [2018] A Time-Domain Preconditioned Truncated Newton Approach to Multiparameter Visco-acoustic Full Waveform Inversion. *SIAM Journal on Scientific Computing*, **40**(4), B1101–B1130.