2018 EAGE ANNUAL
80TH CONFERENCE + EXHIBITION
COPENHAGEN | DENMARK

We A12 04

A Graph-Space Optimal Transport Approach for Full Waveform Inversion

L. Metivier* (CNRS, Univ Grenoble Alpes, LJK/ISTerre), A. Allain (Univ. Grenoble Alpes, LJK), R. Brossier (Univ. Grenoble Alpes, ISTerre), Q. Merigot (Univ. Paris Sud, LMO), E. Oudet (Univ. Grenoble Alpes, LJK), J. Virieux (Univ. Grenoble Alpes, ISTerre)

Summary

The use of optimal transport distances has been recently promoted to mitigate cycle skipping issues in full waveform inversion. This distance is convex with respect to shifted patterns between compared data. This is the reason why it has attracted interest for full waveform inversion as the convexity with respect to time shifts can be seen as a proxy for the convexity with respect to wave velocities. The main difficulty for the application of optimal transport to seismic data is related to their non positivity: the optimal transport theory is developed for the comparison of positive quantities. We present a review of the strategies proposed to overcome this issue, and explain their limitations. They are either not adapted for the comparison of realistic data, or lose the convexity property. On this basis, we propose a novel approach based on a graph space interpretation of the seismic traces. Synthetic and observed traces are seen as point clouds in a 2D space (the graph space), which are compared using a 2D optimal transport technique. This strategy overcomes the positivity issue while preserving the convexity property. An application of this strategy on the Marmousi 2 model illustrates its interest for mitigating cycle skipping.

Introduction

Full waveform inversion (FWI) is a high resolution seismic imaging method based on a data-fitting procedure, which formalism has been introduced in the 80s. This technique is now routinely used both at academy and industry levels, mainly for the reconstruction of subsurface velocities (see Virieux et al., 2017, for instance, for a review). However, successful FWI applications still depend on numerous factors, such as the quality of the initial velocity model, the recording of sufficiently low frequency data, and the ability to design a suitable hierarchical workflow based on specific, possibly combined, data decompositions (frequency, offset, time-window). This lack of flexibility and/or robustness of FWI is mainly associated to what is referred to as cycle skipping or phase ambiguity: the misfit function between synthetic and observed data is not convex with respect to the subsurface parameters. The typical size of FWI problems preventing the use of global optimization algorithms, FWI relies on local optimization solvers which might always converge towards a geologically non-informative local minimum.

Recent attempts to overcome this limitation are based on a misfit function measuring the optimal transport (OT) distance between synthetic and observed data. OT can be seen as a non-local warping problem, where one would try to find the easiest (optimal) mapping between synthetic and observed data. This optimal mapping (transport plan) minimizes the amount of "mass" which needs to be transferred to achieve the mapping, as well as the distance along which it is transferred. The OT distance is convex with respect to shifts between the compared quantities. For FWI, the convexity with respect to time-shifts on seismic traces (or space/time shifts on gathers) can be seen as a proxy for the convexity with respect to wave velocities. This is the reason why OT has attracted interest in the FWI community (Engquist and Froese, 2014; Yang and Engquist, 2017; Métivier et al., 2016a,b; Qiu et al., 2017).

However, OT can not be applied directly to FWI as it is designed for the comparison of positive quantities, while the seismic data is oscillatory. Several propositions have been made to overcome this difficulty. The purpose of this study is first to review these attempts and illustrate some of their limitations. In particular we show that they are either not adapted to FWI or lose the convexity property with respect to time shifts. Then we propose a novel strategy which is adapted to FWI and keeps the convexity property. This strategy is based on the interpretation of seismic traces as point clouds in the graph space. Doing so, we ensure the positivity of the data, and keep the convexity with respect to time-shifts. A numerical experiment on the Marmousi 2 model emphasizes the good properties of the corresponding misfit function to mitigate cycle skipping.

OT for seismic: state-of-the-art

Considering an observed seismic trace $d_{obs}(t)$ and a synthetic trace $d_{cal}(t)$, the p-Wasserstein distance (the OT distance) between d_{obs} and d_{cal} is defined as

$$W_p(d_{obs},d_{cal}) = \left(\min_{\gamma \in \Pi(d_{obs},d_{cal})} \int_0^T \gamma(t,t') ||t-t'||^p \right)^{1/p}, \tag{1}$$

where $\Pi(d_{obs},d_{cal})$ is the ensemble of transport plan $\gamma(t,t')$ such that

$$\Pi(d_{obs},d_{cal}) = \left\{ \gamma(t,t') \geq 0, \ \int_0^T \gamma(t,t')dt = d_{cal}(t'), \ \int_0^T \gamma(t,t')dt' = d_{obs}(t) \right\}. \tag{2}$$

A transport plan $\gamma(t,t')$ describes how much quantity of d_{cal} should be move from t to t' to map d_{cal} to d_{obs}. The OT plan minimizes the sum of all the quantities which are moved multiplied by the distance along which they are moved on the time axis. The OT problem has a solution only if d_{obs} and d_{cal} are positive functions. In addition, the "mass" should be preserved: the time integral of both traces should be the same, such that

$$\int_0^T d_{obs}(t)dt = \int_0^T d_{cal}(t)dt. \tag{3}$$

This property is satisfied for seismic data: this corresponds to the zero frequency, which is always equal to zero. The main difficulty is thus related to the positivity assumption, which is not satisfied.

To overcome this difficulty, different possibilities, yielding different misfit functions, have been explored. An illustration of the convexity of these misfit functions is proposed in Figure 1. The misfit

misfit functions	mass conservation	differentiability	convexity
$f_{+/-}$	✗	✗	✓
f_{scal}	✓	✓	✗
f_{exp}	✗	✓	✓
f_{dual}	✓	✓	✗

Table 1 *Comparison of four strategies to overcome the positivity issue to apply OT to seismic data.*

functions are computed for different time shifts between the observed and synthetic traces, which are taken equal to a Ricker signal. The first strategy consists in comparing separately the positive and negative part of the signal, yielding the misfit function

$$f_{+/-}(m) = W_p^p(d_{obs}^+, d_{cal}^+[m]) + W_p(d_{obs}^-, d_{cal}^-[m]), \tag{4}$$

where $.^+$ and $.^-$ denote positive and negative part operators respectively (Engquist and Froese, 2014). While this ensures the convexity to time shifts (Fig. 1), the resulting misfit function is not differentiable (the positive and negative part operators are not differentiable). In addition, nothing guarantees the mass conservation between the positive and negative part of the observed and synthetic data. A second proposed possibility consists in adding a positive constant c to the observed and synthetic data to make it positive (Yang and Engquist, 2017). To this scaling strategy corresponds the misfit function

$$f_{scal}(m) = W_p^p(d_{obs} + c, d_{cal}[m] + c). \tag{5}$$

This strategy ensures the mass conservation and the differentiability of the misfit function. However, the convexity to time shifts is lost (Fig. 1). A third proposed possibility consists in comparing the exponential of the data, corresponding to the misfit function

$$f_{exp}(m) = W_p^p(\alpha e^{\alpha d_{obs}}, \alpha e^{\alpha d_{cal}[m]}), \tag{6}$$

where α is a user defined positive constant (Qiu et al., 2017). This strategy preserves the convexity to time shifts (Fig. 1) and the differentiability of the misfit function. However, the mass conservation is lost. In addition, the exponential transform distorts the signal, emphasizing large amplitude events at the expense of weaker amplitude events. Finally, another proposed strategy focuses on a specific dual form of the W_1 distance (Métivier et al., 2016a,b), corresponding to the misfit function

$$f_{dual}(m) = \max_{\varphi \in \mathrm{Lip}_1(\mathscr{R})} \int_0^T \varphi(t)\,(d_{cal}[m](t) - d_{obs}(t))\,dt, \tag{7}$$

where $\mathrm{Lip}_1(\mathscr{R})$ is the space of 1-Lipschitz functions

$$\mathrm{Lip}_1(\mathscr{R}) = \left\{\varphi(t),\ \forall (t,t'),\ \left|\varphi(t) - \varphi(t')\right| < \left|t - t'\right|\right\}. \tag{8}$$

This strategy allows to compare non positive data: this maximization problem has a solution even for non positive data. It also preserves the mass conservation and the differentiability of the misfit function. In addition, we have designed an efficient numerical strategy to extend this approach to the comparison of 2D and 3D shot gathers, instead of single traces (Métivier et al., 2016b). However, the convexity to time shifts is lost (Fig. 1). This comparison is summarized in Table 1. As can be seen, no strategy simultaneously satisfies the three criterion: mass conservation, differentiability of the misfit function, convexity with respect to time shifts.

A graph-space approach to optimal transport

To overcome the above-mentioned issues, we propose the graph-space approach. A discretized trace $\mathbf{d} \in \mathscr{R}^N$ is considered as a set of points (t_i, d_i) of $\mathscr{R}^2$. This is the discrete graph of $\mathbf{d}$. The following "smooth" graph space transform is introduced

$$\begin{array}{rccc} \mathscr{G}_\sigma : & \mathbf{d} & \longrightarrow & \mathscr{G}_\sigma(d) = d^{\mathscr{G}_\sigma}(x,t) \\ & \mathscr{R}^N & \longrightarrow & \mathscr{C}^\infty(\mathscr{R}, \mathscr{R}_*^+), \end{array} \tag{9}$$

with

$$d^{\mathscr{G}_\sigma}(x,t) = \frac{1}{2\pi\sigma_x\sigma_t N} \sum_{n=1}^{N} \exp\left(-\frac{(t-t_n)^2}{2\sigma_t^2}\right) \exp\left(-\frac{(x-d_n)^2}{2\sigma_x^2}\right), \tag{10}$$

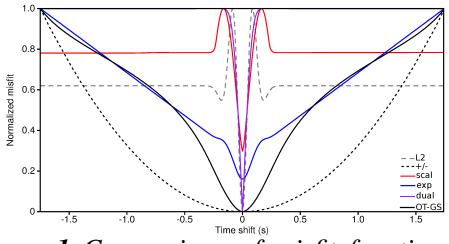

Figure 1 *Comparison of misfit functions depending on time shifts for two shifted in time Ricker signals. Gray dotted line: L^2 misfit function. Black dotted line: $f_{+/-}$. Red solid line: f_{scal}. Blue solid line: f_{exp}. Purple solid line: f_{dual}. Black solid line: f_{OT-GS} with $\tau = 4$.*

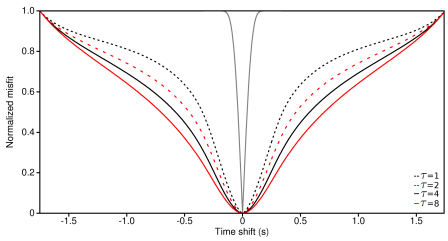

Figure 2 *OT-GS misfit function for decreasing values of τ. Black dotted line: $\tau = 1$. Red dotted line $\tau = 2$. Black solid line $\tau = 4$. Red solid line $\tau = 8$.*

where $\sigma = (\sigma_t, \sigma_x)$ with quantities (σ_t, σ_x) user-defined constant scaling parameters, and $\mathscr{C}^\infty(\mathscr{R}, \mathscr{R}_*^+)$ is the set of strictly positive infinitely differentiable functions of $(\mathscr{R})$. This transform builds a point cloud function from the discrete trace d, where each points is smoothly approximated by a product of exponential functions. This smooth graph transform $\mathscr{G}_\sigma$ preserves the mass conservation: the total mass of the transformed signal $d^{\mathscr{G}_\sigma}$ is equal to 1, and this is true for any discrete trace containing N samples. Based on this transform, we introduce the misfit function

$$f_{OT-GS}(m) = W_p^p\left(\mathscr{G}_\sigma(d_{cal}(m)), \mathscr{G}_\sigma(d_{obs}).\right) \tag{11}$$

The misfit function (11) computes the OT distance between the two "smooth" point clouds associated with the observed and synthetic data. This ensures the mass conservation, the differentiability of the misfit function and preserves the convexity with respect to time shifts, as can be seen on Figure 1. To evaluate numerically f_{OT-GS} and its gradient, 2D OT problems need to be solved. To this purpose, we use the numerical strategy we have introduced in (Métivier et al., 2016b). A scaling factor τ is introduced to control the transport cost along the time axis and the "amplitude" axis in the graph space. Penalizing the transport cost along the amplitude axis favors displacements along the time axis and improves the convexity with respect to time shifts, as can be seen in Figure 2.

Numerical illustration on the Marmousi 2 model

We consider the Marmousi 2 P-wave velocity model downsampled on a 25 m Cartesian grid (Fig. 3a). The water layer is assumed to be known. A fixed-spread surface acquisition at 50 m depth is used with 128 sources located each 130 m from $x = 0.05$ km to $x = 16.7$ km, and 168 receivers, located each 100 m from $x = 0.05$ km to $x = 16.8$ km. The observed data is acquired in the acoustic approximation using a Ricker wavelet centered at 6 Hz and high-pass filtered to remove the energy below 3 Hz. The total recording length is set to 6 s. A free surface condition is implemented at the surface/air interface. We consider a 1D initial model linearly increasing from 1500 m.s^{-1} to 3200 m.s^{-1} (Fig. 3b). We use the l-BFGS algorithm. A Gaussian smoothing of the gradient with a correlation length equal to 0.4 and 0.2 times the local wavelength respectively in horizontal and vertical directions is employed together with a linear depth preconditioning. No stopping criterion is enforced for the l-BFGS algorithm: we let the optimizer free to minimize the misfit function as much as possible. We compare FWI results based on the L^2, f_{dual} applied to 2D shot-gathers, and f_{OT-GS} misfit functions. The figure 2 suggests a hierarchical approach for the OT-GS strategy, with decreasing values of τ. In this example we start with $\tau = 1.5$ and decreases it to $\tau = 0.025$ in 6 stages.

FWI based on the L^2 misfit function fails to recover the P-wave velocity structure: the initial model is marginally updated (Fig. 3c). The main reflectors corresponding to the stronger impedance contrasts appear in the estimated velocity models: however, they are mis-located, due to errors in the large-scale velocity structure. A strong shallow cycle skipping artifact appears under the form of a very low velocity zone at 1 km depth, between $x = 10$ km and $x = 12$ km. FWI based on the misfit function f_{dual} applied to 2D shot-gathers provides slightly better results (Fig. 3d): the central structure between $x = 8$ and $x = 11$ km starts to be recovered, as well as the shallow structure until 1 km depth on the right part of the model. However, the model is not updated below 2.5 km and strong cycle-skipping artifacts can be observed on both sides of the model. FWI using the OT-GS distance yields significantly better results (Fig. 3e,f).

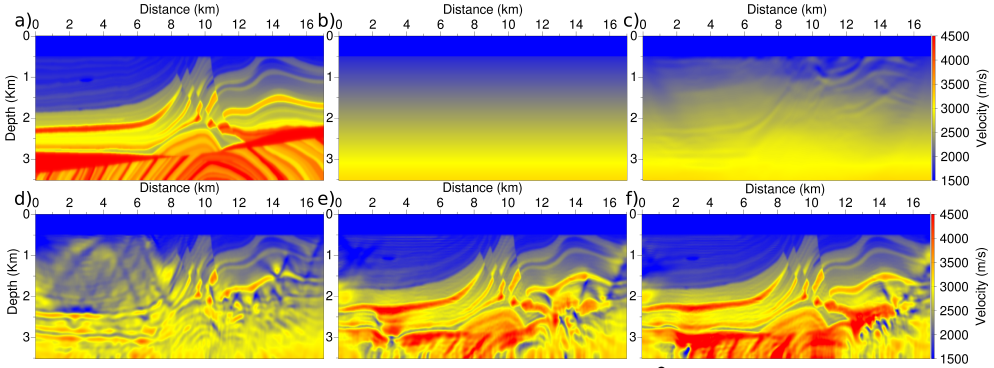

Figure 3 *(a) Exact Marmousi model, (b) 1D initial model, (c) L^2 FWI result, (d) f_{dual} FWI result, (e) OT-GS FWI result with $\tau = 1.5$, (f) OT-GS result final result with $\tau = 0.0025$.*

The velocity structure is already well reconstructed until 3 km depth for $\tau = 1.5$, even if some artifacts can be identified on the edges of the model, especially in the bottom right part of the model (Fig. 3e). Using a hierarchical approach with decreasing values of τ, we are able to enhance the P-wave velocity reconstruction in this zone (Fig. 3f).

Conclusions and perspectives

The graph space approach for using an OT distance in the frame of FWI seems promising. Comparing point clouds overcome the positivity issue while preserving the shape of the signal and the convexity with respect to shifted patterns. The difficulty is now related to the computational cost of the strategy, as a 2D transport problem has to be solved for each trace at each FWI iteration. On the Marmousi example, we have used a decimation in time of a factor 4, together with a rough sampling of the amplitude axis using only 200 points. We also only perform 40 iterations to solve the 2D OT algorithm. In theses settings, the computational time for one gradient is roughly 6 times larger than for the L^2 misfit function. Future work will be devoted to this issue, exploring alternative OT solvers to reduce this computational cost, targeting the implementation of the OT-GS approach for whole shot-gathers instead of individual traces and its application to field data.

Acknowledgements

This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, CHEVRON, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC, STATOIL 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.

References

Engquist, B. and Froese, B.D. [2014] Application of the Wasserstein metric to seismic signals. *Communications in Mathematical Science*, **12**(5), 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 full waveform inversion. *Inverse Problems*, **32**(11), 115008.

Qiu, L., Ramos-Martìnez, J., Valenciano, A., Yang, Y. and Engquist, B. [2017] Full-waveform inversion with an exponentially encoded optimal-transport norm. In: *SEG Technical Program Expanded Abstracts 2017*. 1286–1290.

Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A. and Zhou, W. [2017] An introduction to Full Waveform Inversion. In: Grechka, V. and Wapenaar, K. (Eds.) *Encyclopedia of Exploration Geophysics*, Society of Exploration Geophysics, R1–1–R1–40.

Yang, Y. and Engquist, B. [2017] Analysis of optimal transport and related misfit functions in full-waveform inversion. In: *SEG Technical Program Expanded Abstracts 2017*. 1291–1296.