Increasing the robustness and applicability of full waveform inversion: an optimal transport distance strategy

L. Métivier [1,2], R. Brossier [2], Q. Mérigot [3], E. Oudet [1], J. Virieux [2]
[1] Laboratoire Jean Kuntzmann (LJK), Univ. Grenoble Alpes, CNRS, France
[2] ISTerre, Univ. Grenoble Alpes, France
[3] Laborartoire CEREMADE, Univ. Paris-Dauphine, CNRS, France

7 October 2016

Full waveform inversion starts being used as a standard stage of the seismic imaging workflow, at the exploration scale, for the reconstruction of high resolution wave velocity models. However, its successful application still relies on the estimation of an accurate enough initial velocity model, as well as on the design of a suitable hierarchical workflow, allowing to feed the inversion process progressively with data. These two requirements are mandatory to avoid the cycle skipping or phase ambiguity problem when comparing observed and synthetic data. This difficulty is due to the definition of the full waveform inversion problem as the least-squares minimization of the data misfit. The resulting misfit function has local minima which corresponds to the interpretation of the seismic data up to one or several phase-shifts. In this article, we review an alternative formulation of full waveform inversion based on the optimal transport distance we have proposed in recent studies. We propose to use a particular instance of the optimal transport problem, which is adapted to the interpretation of real seismic data, and for which we design an efficient low-complexity numerical strategy. Numerical results in 2D and 3D configurations (BP 2004, Chevron 2014 benchmark model, SEG/EAGE overthrust model) show that this reformulation should yield a more convex misfit function, less prone to cycle skipping. In this study, we present a simple illustration on the Marmousi model which illustrates how this new distance strongly relaxes the requirement on the initial model design. Starting from a rather simplistic approximation of the initial model, the method is able to reconstruct a meaningful estimation of the Marmousi model, while the standard least-squares formulation is trapped into a local, meaningless minimum.

Key words: High resolution seismic imaging, Inverse problem, Optimal transport, Full Waveform Inversion

1 Introduction

Full waveform inversion (FWI) is a high resolution seismic imaging technique, based on the non-linear iterative minimization of the misfit between calculated and observed data. The simple formalism of the method makes it amenable for the estimation of any subsurface mechanical parameters influencing the wave propagation, such as P- and S- wave velocities, density, attenuation, or anisotropy parameters. Nonetheless, both in industrial and academic applications, FWI is mainly used for the estimation of the wave velocities (Fichtner et al., 2010; Tape et al., 2010; Peter et al., 2011; Sirgue et al., 2010; Plessix and Perkins, 2010; Zhu et al., 2012; Warner et al., 2013; Vigh et al., 2014; Borisov and Singh, 2015; Operto et al., 2015). In addition, the successful application of FWI on real data requires a kinematically compatible initial estimation of the velocity model (computed beforehand through well-established techniques, such as time tomography, velocity analysis, and/or geological interpretation) as well as the design of a specific hierarchical workflow, to interpret the data progressively during iterations, and to converge to a meaningful subsurface model estimation (Virieux and Operto, 2009).

One current trend in FWI is to improve the reconstruction of physical parameters by accounting for a more realistic physics of the wave propagation, including viscosity, elasticity, and anisotropic effects. In particular, this leads to a multi-parameter inversion which should provide much more detailed and reliable characterization of the subsurface, yielding ultimately the possibility to constrain micro-scale parameters through downscaling strategies (Dupuy et al., 2016). Another line of investigation consists in increasing the range of applicability of FWI, mitigating the dependence on ad-hoc initial model build-up, and on the design of a suitable hierarchical workflow. The optimal transport strategy we promote in this study is related to this second line of investigation.

The main reason for the limitation in the applicability of FWI is related to what is usually referred to as cycle skipping, or phase ambiguity. In standard FWI, the oscillatory seismic data is matched in the least-squares sense where each observed sample is compared to the synthetic sample at the same position in time and/or in space. This choice is problematic: if the initial model predicts the signal with a shift larger than half a period, minimizing the least-squares distance between observed and calculated data amounts to match the observed data up to one or several phase shifts. This yields an incorrect estimation of the subsurface model which cannot be overcome through iterations: the optimization is locked into a local minimum. An illustration of this phenomenon, where the seismic data is considered schematically as a sinusoidal temporal signal, is presented in Figure 1. Overcoming this difficulty has been a recurrent objective since the introduction of FWI by Lailly (1983) and Tarantola (1984). Increasing the accuracy of the initial model through high resolution tomography methods, as well as designing hierarchical workflows focusing first on low frequency components of the data, early-arrivals, and/or short offsets, have been initial strategies proposed to challenge this issue (Kolb et al., 1986; Bunks et al., 1995; Pratt, 1999; Shipp and Singh, 2002; Sirgue and Pratt, 2004; Wang and Rao, 2009). They are still the ones implemented for real data applications to guarantee the success of FWI. This careful tuning is case-dependent, therefore, it reduces the flexibility of FWI, and requires an expert usage of FWI and pre-processing tools.

Since the cycle skipping issue results from the use of the least-squares distance to match the seismic data, numerous studies have been proposed to modify this distance within the FWI workflow: cross-correlation of observed and predicted data (Luo and Schuster, 1991), separation of envelope and phase (Fichtner et al., 2008; Bŏzdag et al., 2011), deconvolution approaches (Luo and Sava, 2011; Warner and Guasch, 2014) are instances of such attempts. Images-domain techniques have also been promoted for reflected waves: the velocity accuracy is probed through the analysis of the coherency of migrated images computed in an extended-domain. These images are built upon a generalized image condition, which introduces a fictitious dimension (time-lag or subsurface offsets, for instance) along which the image should be focused at zero if the velocity is correct (see Symes, 2008, for an overview). All these methods are currently under

development: none has yet been proved to overcome definitely the cycle-skipping issue. Also, the extension to 3D framework of these strategies is not always straightforward. This is especially the case for image-domain techniques, due to the high computational cost associated with a higher dimensional model space when repeating the construction of extended-domain images.

In this study, we review an alternative modification of the misfit function, based on the optimal transport distance, we have proposed in [Métivier et al, 2016a,b]. This distance comes from a very active research field in mathematics, as testified by the number of textbooks published on this topic recently (Villani, 2003, 2008; Ambrosio et al., 2008; Santambrogio, 2015). Our proposition is inspired by emerging applications of optimal transport in image processing (see Ferradans et al., 2014; Lellmann et al., 2014, and references therein). In Section 2, we present what is this distance, why its properties make it a suitable choice for increasing the robustness of FWI, and how it can be implemented efficiently, in 2D and 3D configurations, within a standard FWI workflow. In Section 3, we present numerical results on the synthetic Marmousi model emphasizing the interest of this approach. More realistic 2D applications on the BP 2004 salt model and the benchmark 2014 Chevron data are presented in Métivier et al. (2016a), while a 3D application on the SEG/EAGE overthrust model is proposed in Métivier et al. (2016b). Conclusion and perspectives are given in Section 4.

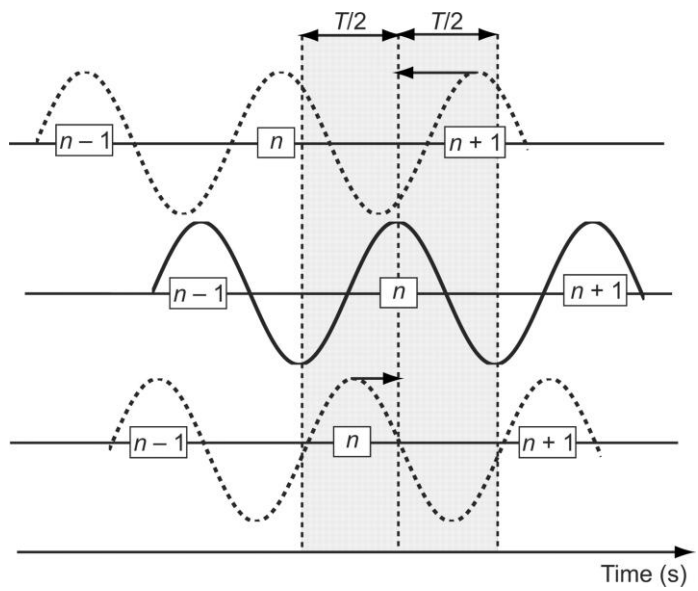

Figure 1: Schematic example of the cycle skipping/phase ambiguity issue on sinusoidal signals. As soon as the initial shift is larger than half a period of the signal, the fit of the signal using a least-squares distance is performed up to one or several phase shifts. On may try to fit the n + 1 dashed wriggle of the top signal with the n continuous wriggle of the middle signal moving to the wrong direction. The bottom dashed signal predicts the n wriggle in less than half-period leading to a correct updating direction (from Virieux and Operto, 2009).

2 Optimal transport for full waveform inversion

2.1 What is optimal transport?

Optimal transport finds its roots back in 1780, in the work of the French engineer Gaspard Monge, from “École des Ponts". Monge supervised a bridge building site. Piles of sand needed to be displaced to fill in holes. Monge expressed mathematically how this displacement could be achieved optimally, to minimize the effort of the workers. This was the first formulation of the optimal transport problem. More than 150 years later, a modern and suitable mathematical formulation for this problem was introduced by Kantorovich (1942), as a minimal relaxation of the Monge problem (when the Monge problem has a solution, the Kantorovich formulation provides this solution). In his description, the initial piles can be associated with N quantities p_i, located at points x_i, for $i = 1, \dots, N$. The configuration of the holes is associated with M quantities q_j, located at points y_j, for $j = 1, \dots, M$. An important assumption is made: the total quantity of sand

requested to fill in the holes q_j is exactly equal to the total quantity of available sand p_i. This corresponds to the mass conservation assumption

$$\sum_{i=1}^{N} p_i = \sum_{j=1}^{M} q_j. \qquad (1)$$

Kantorovich considers the ensemble of displacements making possible to fill the holes with the quantities q_j from the sand piles p_i. These displacements can be represented as matrices with N rows and M columns. An entry γ_{ij} tells how much from the pile p_i should be moved to fill in the hole q_j. Mapping the ensemble of piles p onto the holes q requires that the sum of the elements of the ith row of j is equal to p_i, while the sum of the elements of the jth column of γ_{ij} is equal to q_j. A matrix satisfying this assumption is called a “transport plan". An example of such transport plan is presented in Figure 2. The matrix representing the transport plan schematized in Figure 2 is

$$\gamma = \begin{pmatrix} 3 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 2 & 0 & 0 \end{pmatrix}. \qquad (2)$$

The matrix γ is a transport plan as the sum of the elements of its rows are equal to p_1, p_2 and p_3 respectively, while the sum of the elements of its columns are equal to q_1, q_2, q_3 and q_4 respectively. The interpretation of this transport plan is as follows. From the sand pile p_1, three mass units should be move to q_1, while two mass units should be move to q_3. The sand pile p_2 and p_3 should be moved integrally to q_4 and q_2 respectively.

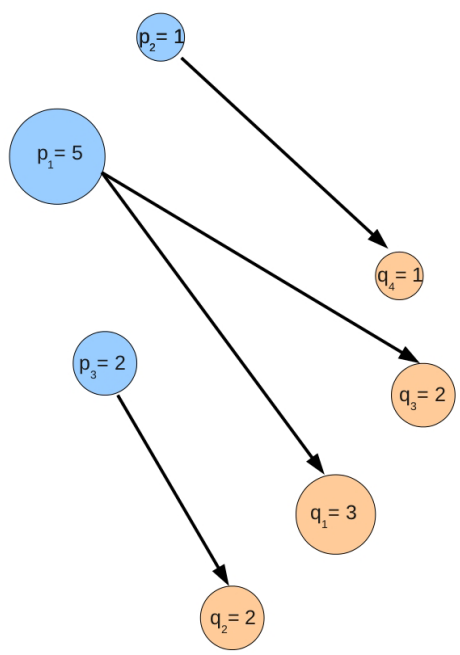

Figure 2: An example of transport plan for the solution of the transport problem between a distribution p_1, p_2, p_3 toward a distribution q_1, q_2, q_3, q_4. We may consider that the distribution p is the discrete observed data and that the distribution q is the discrete synthetic data.

There is an infinity of possible transport plan which allows to move the sand piles p to the holes q. The optimal transport problem consists in computing the unique transport plan which minimizes a function measuring the total displacement cost. This cost is the sum of the elementary costs associated with the elementary displacements. The cost of an elementary displacement between x_i and y_j is measured as the product between the mass which is actually transferred

γ_{ij}, multiplied by the distance between x_i and y_j. This measure implies that a balance has to be found between the amount of mass which is transported, and the distance on which it is transported. Mathematically, this is formulated as the linear programming problem

$$\min_{\gamma_{ij}} \sum_{ij} \gamma_{ij} \|x_i - y_j\|, \text{subject to} \sum_j \gamma_{ij} = p_i, \sum_i \gamma_{ij} = q_j, \qquad (3)$$

where the expression $\|x_i - y_j\|$ denotes a distance between x_i and y_j (often the Euclidean distance).

2.2 Why using an optimal transport distance for full waveform inversion?

The reason why optimal transport can be an interesting tool is that the solution of the linear programming problem (3) defines a distance between the distributions p and q. This distance can be used to measure the discrepancy between any two discrete quantities satisfying the mass conservation assumption (1).

Engquist and Froese (2014) illustrates the interesting behavior of this distance in a simple numerical experiment, where the optimal transport distance between time-shifted Ricker signals is computed. After discretization, two Ricker signals can be considered as two discrete distributions p and q. The distance between the points x_i and y_j where are located masses p_i and q_j is measured in a single dimension, which corresponds to the time axis. As the Ricker signal is not positive, some of the mass p_i and q_j have negative values. While this could be counter-intuitive to define negative mass, it might not be a stringent difficulty for computing the optimal transport distance between the distributions p and q. Engquist and Froese (2014) propose to split the Ricker signals into their positive and negative parts, and to define the optimal transport distance between these signals as the sum of two optimal transport distances: the first computed on the positive part of the Ricker signals, the second computed on the opposite of their negative parts. Working with the dual formulation of the optimal transport distance (which is defined in the next section), yields also the possibility to consider directly positive and negative mass. This is the strategy we employ. Finally, as the Ricker signals are only shifted in time, the distribution p and q satisfy the mass conservation assumption (1). Therefore, the optimal transport distance between two shifted in time Ricker signals can be computed.

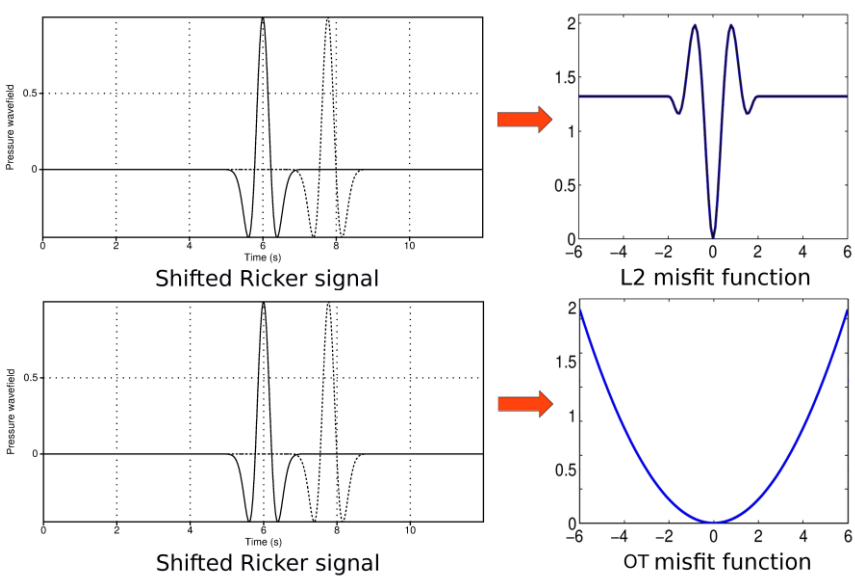

Figure 3: Computation of the misfit function between two time-shifted Ricker signal depending on the time shift, using a least-squares distance and an optimal transport distance. While the least-squares distance yields a non-convex misfit function with two local minima aside the global minimum at zero time-shift, the optimal transport distance yields a perfectly convex misfit function (from Engquist and Froese, 2014).

When a standard least-squares norm is used, the distance between the Ricker signals with respect to the time shift presents two local minima aside the global minimum. This is an illustration of cycle skipping: minimizing the misfit between these two signal starting from a too large time shift would result in a misinterpretation of the signal. Conversely, when the optimal transport distance is used, the distance with respect to the time shift is perfectly convex: a single minimum exists. This is illustrated in Figure 3.

How can we interpret this result? The definition of optimal transport is the key: the distance between the two signals is the minimum effort required to transport one signal onto the other. This effort depends on the distance between points from which the mass has to be exchanged. Therefore, as soon as the time shift increases, the transportation effort increases.

2.3 What are the difficulties for using an optimal transport distance for full waveform inversion?

Using an optimal transport distance, beyond simple Ricker signals, for real seismic data, is not straightforward. The first main difficulty is related to the mass conservation assumption (1). This assumption cannot be guaranteed for real seismic data. Indeed, there is no reason for the mass distributions coming from the discretization of observed and calculated seismic traces to satisfy this assumption. Seismic events recorded in the observed data may not be predicted by the calculated data, resulting in this disparity. This is a serious difficulty. If this assumption is not satisfied, the Kantorovich problem has no solution. In the context of exploration scale real seismic data, the important density of sources and receivers, yielding large number of seismic traces, also raises the question how the optimal transport should be used to measure the misfit between observed and calculated data. A straightforward use consists in considering each seismic trace independently and solve as many 1D optimal transport problem as the number of traces. In this framework, the resulting misfit functions should be the sum of the optimal transport misfit computing for each trace.

However, optimal transport yields the possibility to perform more interesting multi-dimensional comparisons. Seismic imaging practitioners have used over year and year representations of the seismic data in panels allowing to identify clearly the type of waves and the seismic events which have been recorded. The identification of these events is performed through the analysis of their coherency in receiver-gather or shot-gather panels. Interestingly, wave velocity perturbations are responsible in this case for shifting these events not only in the time coordinate, but also in the receiver or shot coordinate. Accounting robustly for the shifts between these events can thus be achieved by defining an optimal transport distance in a multi-dimensional space: for time-domain FWI, in 2D configurations, 2D transport problems should be considered (time-axis + 1D receiver array), while in 3D configurations, 3D transport problems should be considered (time-axis + 2D receiver network). An additional dimension along the source dimension could be also considered. However, this would break the standard parallelism on shots commonly used for time-domain FWI implementations. For frequency-domain FWI, however, in 2D configuration, 2D transport problems should be considered (1D source array+ 1D receiver array), while in 3D, 4D transport problems could be considered: (2D source network+ 2D receiver network). Let us note that, for frequency-domain application, the data has to be split into its real part and imaginary part: the resulting distance should be computed as the sum of the optimal transport distance computed separately on these two quantities. This ability to account for the coherency of the seismic events not only in the time dimension but in the whole gather domain is a key feature of the optimal transport distance.

In practice, the discretization of 2D or 3D realistic scale shot gathers yields mass distributions that can reach several tenths of millions of elements and beyond. This raises another important difficulty. Standard algorithms for solving linear programming problems such as the problem (3) have a high computational cost, not adapted to this problem scale. For instance, all

known methods for solving the problem (3) exactly, assuming the data are described by integers, have a computational complexity large than $\mathcal{O}(N^2)$, where the total number of discrete samples representing a gather is denoted by N. For realistic scale applications, algorithms in linear complexity $\mathcal{O}(N)$ or quasilinear complexity $\mathcal{O}(NlogN)$ are almost mandatory.

Our contribution is the design of a method which is able to overcome these difficulties of seismic imaging; namely a strategy to estimate an optimal transport distance between mass distribution which do not satisfy the mass conservation assumption, with a computational complexity at most quasilinear.

2.4 What is the solution we propose?

The method we propose consists in solving a modified version of the Kantorovich problem. This strategy is explained in details in either an article oriented to the geophysical community (Métivier et al., 2016a) or an article oriented to the applied-mathematics community (Métivier et al., 2016b). Only the main ideas, on which this strategy is based, are presented here in order to identify salient features of this new distance definition. While the problem (3) is referred to as the primal problem, the associated dual problem can be expressed as finding values φ_i at each discrete point of the dataset subject to inequality constraints:

$$\max_{\varphi_i} \sum_i \varphi_i\left(d_{obs,i} - d_{cal,i}\right), \qquad \left|\varphi_i - \varphi_j\right| \leq \left\|x_i - x_j\right\|, \forall i,j = 1,\dots,N. \tag{4}$$

Please note that we compute the difference of observed and calculated values at the same point i while the inequality constraint is non-local with the two indexes i and j. The solution of this dual problem (4) is equivalent to the solution of the primal problem (3). In other words, the distance between the observed data d_{obs} (corresponding to the previously introduced distribution p) and the data d_{cal} (corresponding to the previously introduced distribution q) computed as solution of these two problems is the same. This very important duality result is due to Kantorovich (a complete proof can be found in Villani (2003) or Santambrogio (2015)). Exactly as for the primal problem, if the mass conservation assumption (1) is not satisfied, the dual problem has no solution.

Our first contribution is to recognize that, however, this can be relaxed through the addition of the constraints $|\varphi_i| \leq c$ into the dual problem, where the quantity c is a user defined constant quantity. The modified dual transport problem we consider is thus

$$\max_{\varphi_i} \sum_i \varphi_i\left(d_{obs,i} - d_{cal,i}\right), \left|\varphi_i - \varphi_j\right| \leq \left\|x_i - x_j\right\|, \forall i,j = 1,\dots,N, |\varphi_i| \leq c, \forall i = 1,\dots,N. \tag{5}$$

This dual transport problem accepts a solution when the mass conservation assumption (1) is not satisfied. The problem (5) defines the Kantorovich-Rubinstein norm $\|d_{obs} - d_{cal}\|_{KR}$ (Lellmann et al., 2014).

As a second step, we focus on the particular case where, instead of the Euclidean distance $\left\|x_i - x_j\right\|$, we use the ℓ_1 distance we denote by $\left|x_i - x_j\right|$ to measure the distance between point x_i and x_j. For a 2D dataset, if we denote $x_i^1$and x_i^2 the two components of the point x_i, we would have

$$\left\|x_i - x_j\right\| = \sqrt{\left(x_i^1 - x_j^1\right)^2 + \left(x_i^2 - x_j^2\right)^2}, \left|x_i - x_j\right| = \left|x_i^1 - x_j^1\right| + \left|x_i^2 - x_j^2\right|. \tag{6}$$

The transport problem we consider is thus

$$\max_{\varphi_i} \sum_i \varphi_i\left(d_{obs,i} - d_{cal,i}\right), \left|\varphi_i - \varphi_j\right| \leq \left|x_i - x_j\right|, \forall i,j = 1,\dots,N, |\varphi_i| \leq c, \forall i = 1,\dots,N. \tag{7}$$

The complexity of the transport problem (7) remains important: even if the size of the variable φ is N, the number of linear inequalities constraints to be satisfied is in $\mathcal{O}(N^2)$. The interest for focusing on the Kantorovich-Rubinstein norm associated with the ℓ_1 distance is related to the

fact that an equivalent formulation can be derived, involving only $\mathcal{O}(N)$ linear constraints. This drastic reduction comes from a particular feature of the ℓ_1 distance (also known as Manhattan distance), which allows to satisfy the linear constraints only locally. Satisfying these $\mathcal{O}(N)$ local constraints is equivalent to satisfy the $\mathcal{O}(N^2)$ global constraints. The following equivalent formulation can thus be deduced from (7), expressed as

$$\max_{\varphi_i} \sum_i \varphi_i\left(d_{obs,i} - d_{cal,i}\right), |\varphi_i - \varphi_{i+1}| \leq |x_i - x_{i+1}|, \forall i = 1, \dots, N, |\varphi_i| \leq c, \forall i = 1, \dots, N. \quad (8)$$

The problem (8) is a modified dual optimal transport problem, which has a solution even when the mass conservation assumption (1) is not satisfied, and which can be expressed as a linear programming problem of size N with $\mathcal{O}(N)$ linear inequality constraints. This reformulation opens the way to the design of efficient numerical strategies for its solution. To this purpose, the main idea we exploit is to recast it as a non-smooth optimization problem, which can be solved efficiently through a proximal splitting algorithm named SDMM (Combettes and Pesquet, 2011). Using this method, we design an algorithm for approximating the solution of (8) with linear complexity $\mathcal{O}(N)$ or quasilinear complexity $\mathcal{O}(NlogN)$, allowing to consider the large-scale problems imposed by full waveform inversion, both for 2D and 3D data panels associated with 2D and 3D configurations. The full procedure is described in Métivier et al. (2016b)

An illustration of the use of this optimal transport distance on the example of the shifted in time Ricker signal is provided in Figure 4. Compared to the least-squares distance, a single minimum is recovered. However, compared to the standard optimal transport distance used by Engquist and Froese (2014), the convexity of the misfit function with respect to the time shift is lost. This is probably due to the use of the ℓ_1 norm to measure the spatial distance between points, even if this remains to be proved theoretically. This loss of convexity could appear as a penalizing feature; however, for the practical applications we have considered, this does not seem to have an impact on the convergence of the local optimization solvers.

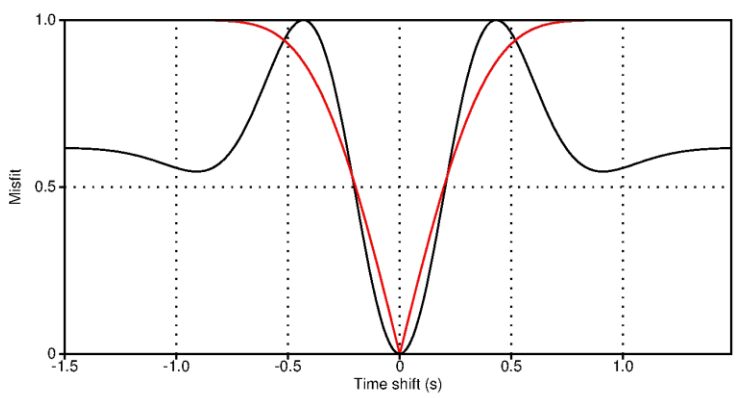

Figure 4: Computation of the misfit function between two time-shifted Ricker signal depending on the time shift, using a least-squares distance (black) and the Kantorovich-Rubinstein distance (red). We recover a single minimum. However, compared to the optimal transport distance used by Engquist and Froese (2014), the convexity of the misfit function is lost at the minimum. This is due to the particular formulation of the Kantorovich Rubinstein problem which is based on a ℓ_1 measure of the distance between points.

3 Implementation and numerical results

3.1 Implementation: gradient computation

The FWI problem is solved through local optimization techniques, which are based on the computation of the gradient of the misfit function. Modifying the misfit function thus requires to

examine how to compute its gradient. Standard implementation of FWI are based on the adjoint-state approach: the gradient is computed as the zero-lag correlation between incident and adjoint wavefields. The adjoint wavefields are computed as the backpropagation of residuals at the receivers location (Plessix, 2006). It is established that a modification of the misfit function only induces a modification of the adjoint source definition (see Brossier et al. (2010); Luo and Sava (2011) for instance). The question for us is thus to determine what are the residuals associated with the optimal transport misfit function (8) which will act as adjoint source.

The computation of the misfit function requires the solution of the problem (8). The misfit function is the value of the quantity $\sum_i \bar{\varphi}_i\left(d_{obs,i} - d_{cal,i}\right)$ where the function φ maximizes this quantity under the constraints which have been discussed earlier. The important result regarding the implementation of our strategy is the following: it is possible to show that the residuals associated with this misfit function is precisely the function $\bar{\varphi}$. Therefore, the optimal transport problem has to be solved only once by iteration, which prevents drastic increase in computer time (for instance we observe a 20% increase of the computation time for the gradient, on the Marmousi example presented after). From its solution, we can extract both the value of the misfit function and the residuals which have to be back-propagated for computing the adjoint fields and, therefore, building the model gradient.

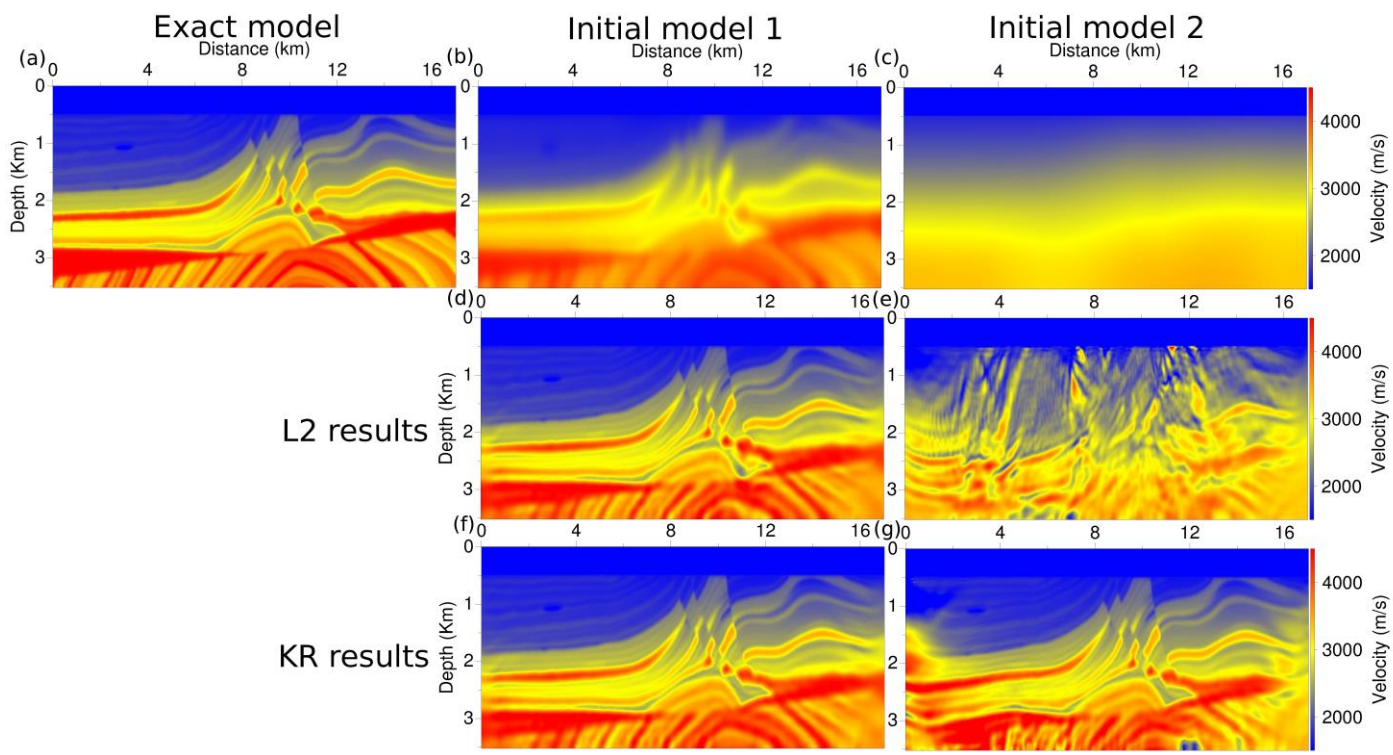

Figure 5: Marmousi model case study. Exact model (a), initial model 1 (b), initial model 2 (c), results obtained with the least-squares distance starting from model 1 (d), from model 2 (e), results obtained with the KR distance starting from model 1 (f), from model 2 (g).

3.2 Numerical results: Marmousi case study

We illustrate the Kantorovich-Rubinstein approach on the following Marmousi synthetic case study. The exact Marmousi model is presented in Figure 5(a). A synthetic dataset is computed in the 2D acoustic constant-density approximation. We use a fixed-spread surface

acquisition with 128 sources each 125 m and 168 receivers each 100 m, at 50 m depth. A Ricker source function centered on 5 Hz is used to generate the synthetic dataset. The frequency content of the source is high-pass filtered above 3 Hz to mimic realistic seismic data. Below 3 Hz, seismic signal is contaminated by noise, and therefore unavailable for inversion. Two initial models are considered: the first contains the main features of the exact model, only with smoother interfaces (Fig. 5.b). The second is a strongly smoothed version of the exact model with very weak lateral variations and underestimated growth of the velocity in depth (Fig. 5.c). Starting from these two initial models, we compare the FWI results obtained using a least-squares distance and the optimal transport distance we propose. The minimization is performed using the l-BFGS algorithm (Nocedal, 1980) implemented in the SEISCOPE optimization toolbox (Métivier and Brossier, 2016).

The results we obtain are presented in Figure 5(d-g). The convergence to a correct estimation of the P-wave velocity model is obtained using both the least-squares (Fig. 5.d) and optimal transport (Fig. 5.f) distances starting from the first initial model. A slightly better estimation of the low velocity zone at x = 11 km, z = 2.5 km is obtained using the optimal transport distance. A high velocity artifact can be seen for the least-squares estimation in this zone. More importantly, starting from the second initial model, only the results obtained using the optimal transport distance are meaningful (Fig.5g). The poor initial approximation of the P-wave velocity is responsible for cycle skipping and the least-squares estimation converges towards a local minimum (Fig.5f). The estimation obtained with the optimal transport distance is significantly closer from the true model, despite low velocity artifacts in the shallow part at x = 1.5 km, z = 1 km and in depth at x = 12 km, z = 3.4 km. This example illustrates the potential of optimal transport for FWI: starting from a very crude approximation of the P-wave velocity, a meaningful estimation is computed. In the same configuration, FWI based on the least-squares distance fails and produces a heavily cycle skipped estimation.

4 Conclusion

Optimal transport distance appears as an interesting tool for FWI, for its ability to provide a more convex misfit function. In addition, contrary to other misfit function modifications such as cross-correlation techniques, the resolution power seems to be preserved. Its principle is based on a change in the way the distance should be computed between observed and computed seismograms. When a least-squares distance is used to compare the seismic data, only point-to-point comparisons are performed and the information on the coherency between seismic events in the shot-gather panel is not accounted for in the inversion. Optimal transport offers the possibility to account for this coherency in the inversion. Theoretical work is mandatory to better understand the more convex behavior of the modified optimal transport distance we promote. However, this ability to take into account this coherency of the data is the reason why we consider it as a promising alternative to standard misfit functions for FWI.

Currently, the method has been successfully applied for a salt reconstruction problem on the 2D BP2004 model. Starting from an initial model containing no information about the presence of salt, we were able to design a suitable workflow to progressively reconstruct a complex salt body. This prompts us to continue the investigation of this strategy for imaging subsalt targets. The optimal transport method has also been successfully applied to the Chevron 2014 benchmark dataset, for which results comparable with those published in the literature have been recovered, using a very simple workflow. These two applications are presented in Métivier et al. (2016a). A first 3D application on the SEG/EAGE overthrust model is also presented in Métivier et al. (2016b). Future work will now include applications on 2D and 3D real data, for instance on the Valhall dataset. Methodological work is also on going, as other optimal transport formulation might improve further the method in terms of computational cost and misfit function convexity.

5 Acknowledgments

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

References

- Ambrosio, L., Gigli, N., and Savaré, G. (2008). Gradient flows: in metric spaces and in the space of probability measures. Springer Science & Business Media.
- Borisov, D. and Singh, S. C. (2015). Three-dimensional elastic full waveform inversion in a marine environment using multicomponent ocean-bottom cables: a synthetic study. Geophysical Journal International, 201:1215-1234.
- Brossier, R., Operto, S., and Virieux, J. (2010). Which data residual norm for robust elastic frequency domain full waveform inversion? Geophysics, 75(3):R37-R46.
- Bunks, C., Salek, F. M., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. Geophysics, 60(5):1457-1473.
- Bŏzdag, E., Trampert, J., and Tromp, J. (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. Geophysical Journal International, 185(2):845-870.
- Combettes, P. L. and Pesquet, J.-C. (2011). Proximal splitting methods in signal processing. In Bauschke, H. H., Burachik, R. S., Combettes, P. L., Elser, V., Luke, D. R., and Wolkowicz, H., editors, Fixed-Point Algorithms for Inverse Problems in Science and Engineering, volume 49 of Springer Optimization and Its Applications, pages 185-212. Springer New York.
- Dupuy, B., Asnaashari, A., Brossier, R., Garambois, S., Métivier, L., Ribodetti, A., and Virieux, J. (2016). A downscaling strategy from FWI to microscale reservoir properties from high-resolution images. The Leading Edge, 35:1146-1150.
- Engquist, B. and Froese, B. D. (2014). Application of the wasserstein metric to seismic signals. Communications in Mathematical Science, 12(5):979-988.
- Ferradans, S., Papadakis, N., Peyré, G., and Aujol, J.-F. (2014). Regularized Discrete Optimal Transport. SIAM Journal on Imaging Sciences, 7(3):1853-1882.
- Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. P. (2008). Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain. Geophysical Journal International, 175:665-685.
- Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. P. (2010). Full waveform tomography for radially anisotropic structure: New insights into present and past states of the Australasian upper mantle. Earth and Planetary Science Lettters, 290(3-4):270-280.
- Jannane, M., Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., Singh, S., Snieder, R., Tarantola, A., and Trezeguet, D. (1989). Wavelengths of Earth structures that can be resolved from seismic reflection data. Geophysics, 54(7):906-910.
- Kantorovich, L. (1942). On the transfer of masses. Dokl. Acad. Nauk. USSR, 37:7-8.
- Kolb, P., Collino, F., and Lailly, P. (1986). Prestack inversion of 1-D medium,. In Extended Abstracts, 74, pages 498-508.
- Lailly, P. (1983). The seismic problem as a sequence of before-stack migrations. In Bednar, J., editor, Conference on Inverse Scattering: Theory and Applications. SIAM, Philadelphia.
- Lellmann, J., Lorenz, D., Schöonlieb, C., and Valkonen, T. (2014). Imaging with

kantorovich-rubinstein discrepancy. SIAM Journal on Imaging Sciences, 7(4):2833-2859.
- Luo, S. and Sava, P. (2011). A deconvolution-based objective function for wave-equation inversion. SEG Technical Program Expanded Abstracts, 30(1):2788-2792.
- Luo, Y. and Schuster, G. T. (1991). Wave-equation traveltime inversion. Geophysics, 56(5):645-653.
- Métivier, L. and Brossier, R. (2016). The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication. Geophysics, 81(2):F11-F25.
- 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, September 2016.
- Nocedal, J. (1980). Updating Quasi-Newton Matrices With Limited Storage. Mathematics of Computation, 35(151):773-782.
- Operto, S., Miniussi, A., Brossier, R., Combe, L., Méetivier, L., Monteiller, V., Ribodetti, A., and Virieux, J. (2015). Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation. Geophysical Journal International, 202(2):1362-1391.
- Peter, D., Komatitsch, D., Luo, Y., Martin, R., Le Go, N., Casarotti, E., Le Loher, P., Magnoni, F., Liu, Q., Blitz, C., Nissen-Meyer, T., Basini, P., and Tromp, J. (2011). Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes. Geophysical Journal International, 186(2):721-739.
- 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.
- Plessix, R. E. and Perkins, C. (2010). Full waveform inversion of a deep water ocean bottom seismometer dataset. First Break, 28:71-78.
- Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part I : theory and verification in a physical scale model. Geophysics, 64:888-901.
- Santambrogio, F. (2015). Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling. Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing.
- Shipp, R. M. and Singh, S. C. (2002). Two-dimensional full wave eld inversion of wide-aperture marine seismic streamer data. Geophysical Journal International, 151:325-344.
- Sirgue, L., Barkved, O. I., Dellinger, J., Etgen, J., Albertin, U., and Kommedal, J. H. (2010). Full waveform inversion: the next leap forward in imaging at Valhall. First Break, 28:65-70.
- Sirgue, L. and Pratt, R. G. (2004). Effcient waveform inversion and imaging : a strategy for selecting temporal frequencies. Geophysics, 69(1):231-248.
- Symes, W. W. (2008). Migration velocity analysis and waveform inversion. Geophysical Prospecting, 56:765-790.
- Tape, C., Liu, Q., Maggi, A., and Tromp, J. (2010). Seismic tomography of the southern California crust based on spectral-element and adjoint methods. Geophysical Journal International, 180:433-462.
- Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation.

Geophysics, 49(8):1259-1266.
- Vigh, D., Jiao, K., Watts, D., and Sun, D. (2014). Elastic full-waveform inversion application using multicomponent measurements of seismic data collection. Geophysics, 79(2):R63-R77.
- Villani, C. (2003). Topics in optimal transportation. Graduate Studies In Mathematics, Vol. 50, AMS.
- Villani, C. (2008). Optimal transport : old and new. Grundlehren der mathematischen Wissenschaften.Springer, Berlin.
- Virieux, J. and Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. Geophysics, 74(6):WCC1-WCC26.
- Wang, Y. and Rao, Y. (2009). Reflection seismic waveform tomography. Journal of Geophysical Research, 114(B03304):doi:10.1029/2008JB005916.
- Warner, M. and Guasch, L. (2014). Adaptative waveform inversion - FWI without cycle skipping - theory. In 76th EAGE Conference and Exhibition 2014, page We E106 13.
- Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., Stekl, I., Guasch, L., Win, C., Conroy, G., and Bertrand, A. (2013). Anisotropic 3D full-waveform inversion. Geophysics, 78(2):R59-R80.
- Zhu, H., Bŏzdag, E., Peter, D., and Tromp, J. (2012). Structure of the european upper mantle revealed by adjoint tomography. Nature Geoscience, 5:493-498.