

# An optimal transport distance for full waveform inversion: application to the 2014 Chevron benchmark data-set

L. Métivier\*, LJK, CNRS, Univ. Grenoble Alpes, R. Brossier, ISTERre, Univ. Grenoble Alpes, Q. Mériqot, CEREMADE, CNRS, Univ. Paris Dauphine, E. Oudet, LJK, Univ. Grenoble Alpes, J. Virieux, ISTERre, Univ. Grenoble Alpes

## SUMMARY

Mitigating cycle skipping in full waveform inversion is a long term issue. In this study, a modification of the misfit function based on an optimal transport distance is proposed. A specific numerical strategy based on the Kantorovich-Rubinstein norm and a proximal splitting algorithm is designed to compute this misfit function and its gradient efficiently. An example of application on the Marmousi model shows that the optimal transport misfit function is more robust to cycle skipping than the conventional  $L^2$  misfit function. When applied to the Chevron 2014 benchmark data-set, a reliable velocity estimation is computed, using only a multi-scale strategy in frequency, while the  $L^2$  misfit function is trapped into a local minimum following the same workflow.

## INTRODUCTION

Full waveform inversion is a high resolution subsurface imaging technique, based on the minimization of the distance between observed and calculated data (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009). Despite the formalism of the method allows for the reconstruction of any parameter influencing the wave propagation (P- and S-wave velocities, density, attenuation, anisotropy parameters), the method is mainly used as a high resolution velocity model building tool (Plessix and Perkins, 2010; Warner et al., 2013; Operto et al., 2015). The large-scale aspect of the problem leads to using local optimization techniques based on the gradient direction rather than global optimization techniques relying on an exhaustive exploration of the model space.

As large-to-intermediate wavelength velocity perturbations have mainly a kinematic impact on the wave propagation (Jannane et al., 1989), synthetic data computed in a velocity model lacking this low wavenumber information exhibit time-shifted patterns compared to the observed data. From an inverse problem point of view, recovering this low wavenumber information thus requires to correctly account for the time-shifts between observed and recorded data. Minimizing the  $L^2$  distance between observed and synthetic data using local optimization techniques is not the best choice to achieve this goal: unless the time-shifts are lower than half the dominant period of the signal in the initial model, local optimization techniques will converge to a local minimum corresponding to a model predicting the data with one or several phase shifts. This difficulty is referred to as cycle skipping.

Overcoming cycle skipping has been the topic of numerous studies, following three main lines of investigation. The first consists in a hierarchical/multi-scale interpretation of the data. As the dominant phase is larger for low frequency data, the valley of attraction of the  $L^2$  misfit function is broadened, thus relaxing the constraint on the initial model accuracy (Bunks et al., 1995; Pratt, 1999). The second consists in modifying

the misfit function in an attempt to design a convex function with respect to the velocity model. Cross-correlation (Luo and Schuster, 1991), warping (Ma and Hale, 2013) and deconvolution approaches (Luo and Sava, 2011; Warner and Guasch, 2014) belong to this class of methods. All are based on an (automatic) extraction of the time-shifts between single traces, and a weighting function aiming at refocusing the energy at zero time-shift. The third category of methods (applicable only to reflection data) is referred to as extended image domain technique (Symes, 2008). The accuracy of the velocity model is probed through the construction of reflectivity images using an extended image condition (angle, subsurface offset, time-lag). The misfit function is defined to refocus the energy at zero in the extended dimension.

More recently, Engquist and Froese (2014) have proposed to make use of an optimal transport distance for comparing seismic data. Optimal transport is an emerging tool in image processing having its roots in the mathematical work of Monge (1781) and more recently Kantorovich (1942). Optimal transport can be seen as the computation of maps allowing to transfer mass (here data values) from an initial configuration to a final configuration with a minimum effort. As such, it defines a distance between these two configurations, corresponding to this minimum effort. Considering image values as a mass distribution (the gray level of each pixel corresponding to a given amount of mass), optimal transport gives a tool to perform global comparisons between images (as opposed to local, pixel-by-pixel comparisons, performed by  $L^p$  norms), which can be used for pattern recognition for instance. This ability to perform patterns recognition in two images makes it an interesting tool for seismic imaging as it should be able to detect time-shifts between seismograms. This intuition is confirmed for 1D signals by Engquist and Froese (2014) where the optimal transport distance between shifted Ricker signals is shown to be convex with respect to time-shifts between these signals.

Extending the use of the optimal transport distance to industrial scale seismic data-sets faces two main difficulties. The first is related to the conservation of the mass between the compared distribution: standard optimal transport problems are based on the assumption that the total mass is conserved from the initial configuration to the final configuration. This is not a reasonable assumption for observed and calculated seismic data. Missing events in the calculated data (for instance reflections) lead to a lack of mass in the initial configuration. The second difficulty is related to the computation cost. Industrial data-sets involve several thousands time-steps and receivers leading to optimal transport problems with complexity from  $O(10^6)$  in 2D to  $O(10^9)$  in 3D. A drastic simplification can be obtained by considering each seismic trace independently (therefore reducing the complexity to a collection of 1D optimal transport problems). However this implies not accounting for the lateral coherency of the seismic events when

comparing data; Preserving this information in the data comparison might be beneficial (if not essential) to the inversion.

In this study, a strategy to overcome these two difficulties is proposed. The optimal transport distance which is introduced is the one associated with the Kantorovich-Rubinstein (KR) norm which allows for non- conservation of the mass (Lellmann et al., 2014). An original numerical method is designed to approximate the solution of the optimal transport problem in linear complexity, therefore allowing to tackle the large scale problems associated with industrial scale data-sets. The interest of the method is first presented on the Marmousi case, then it is applied to the Chevron 2014 synthetic benchmark data-set, showing in both cases a significant enhancement of the velocity reconstruction compared to the results obtained with a  $L^2$  distance.

## METHOD

FWI is usually formulated as the nonlinear least-squares minimization problem

$$\min_m f_{L^2}(m) = \frac{1}{2} \sum_{s=1}^{N_s} \|d_{pred,s}[m] - d_{obs,s}\|^2, \quad (1)$$

where  $m$  is a set of subsurface parameter,  $N_s$  the number of sources,  $\|\cdot\|$  denotes the  $L^2$  norm,  $d_{obs,s}$  is the observed data and  $d_{pred,s}[m]$  is the data predicted through the solution of the forward problem

$$A(m)u_s = s, \quad d_{pred,s}[m] = u_s(x_r, t), \quad r = 1, \dots, N_r. \quad (2)$$

In equation (2),  $A(m)$  is a general wave equation operator (from acoustics to visco-anisotropic-elastodynamics). In this study, the usual  $L^2$  norm is replaced with the KR norm such that the FWI problem becomes

$$\min_m f_{KR}(m) = \sum_{s=1}^{N_s} \|d_{pred,s}[m] - d_{obs,s}\|_{KR}. \quad (3)$$

For the sake of simplicity, we assume  $N_s = 1$  in the following and drop the index  $s$ . Extension to  $N_s > 1$  is straightforward by summation. The computation of the KR norm requires the solution of the maximization problem

$$\|d_{pred}[m] - d_{obs}\|_{KR} = \max_{\varphi \in BL_1} \int_{x_r} \int_0^T \varphi(x_r, t) (d_{pred}[m](x_r, t) - d_{obs}(x_r, t)) dx_r dt, \quad (4)$$

where  $BL_1$  denotes the space of bounded 1-Lipschitz functions for the  $L^1$  distance such that

$$\begin{aligned} \forall(x, t), \quad & |\varphi(x, t)| < 1, \\ \forall(x, t), (x', t'), \quad & |\varphi(x, t) - \varphi(x', t')| \leq |x - x'| + |t - t'|. \end{aligned} \quad (5)$$

In discrete form, this amounts to solve the linear programming problem

$$\begin{aligned} \max_{\varphi} g(\varphi) &= \sum_{n=1}^{N_t} \sum_{i=1}^{N_r} \varphi_{ni} (d_{pred}[m]_{ni} - (d_{obs})_{ni}) \Delta x \Delta t \\ \forall(n, i), \quad & |\varphi_{ni}| < 1 \\ \forall(n, i), (n', i'), \quad & |\varphi_{ni} - \varphi_{n'i'}| \leq |t_n - t_{n'}| + |x_i - x_{i'}|. \end{aligned} \quad (6)$$

In Métivier et al. (2016), we show that the complexity of this problem can be reduced from  $O(N^2)$  to  $O(N)$  constraints by

deriving an equivalent problem using only local constraints (thanks to a property of the  $L^1$  norm). This equivalent problem is

$$\begin{aligned} \max_{\varphi} g(\varphi) &= \sum_{n=1}^{N_t} \sum_{i=1}^{N_r} \varphi_{ni} (d_{pred}[m]_{ni} - (d_{obs})_{ni}) \Delta x \Delta t \\ \forall(n, i), \quad & |\varphi_{ni}| < 1 \\ \forall(n, i), \quad & |\varphi_{n+1i} - \varphi_{ni}| \leq |t_{n+1} - t_n| \\ \forall(n, i), \quad & |\varphi_{ni+1} - \varphi_{ni}| \leq |x_{i+1} - x_i|. \end{aligned} \quad (7)$$

The solution of this problem is obtained by recasting it as the convex non-smooth problem

$$\max_{\varphi} g(\varphi) + i_K(C\varphi), \quad (8)$$

where  $K$  is the unit hypercube  $\{x \in \mathbb{R}^P, |x_i| < 1\}$  with  $P$  the total number of discrete constraints,  $i_K$  the indicator function of  $K$  such that

$$i_K = \begin{cases} 0 & \text{if } x \in K \\ -\infty & \text{if } x \notin K, \end{cases} \quad (9)$$

and  $C$  is the matrix encoding the local linear constraints, such that

$$C = [D_t \ D_{x_r} \ I]^T, \quad (10)$$

where  $D_t$  and  $D_{x_r}$  are the forward finite difference operators in the time and receiver dimensions respectively (local Lipschitz constraints) and  $I$  is the identity matrix (absolute value constraints). The solution of the problem (8) is obtained using the proximal splitting algorithm Simultaneous Direction Method of Multipliers (SDMM) (Combettes and Pesquet, 2011). This algorithm only requires the computation of the proximal operators of  $g(\varphi)$  and  $i_K$  for which closed-form expression can be obtained, and which can be evaluated in linear complexity. The SDMM algorithm also requires at each iteration the solution of a linear system involving the operator  $C^T C = I + D_{x_r}^T D_{x_r} + D_t^T D_t$  which can be proved to be the second-order finite-difference discretization of the 2D Laplacian operator with Neumann homogeneous boundary conditions. Fast solvers for the solution of such linear systems can thus be used, either relying on the Fast Fourier Transform (Swarztrauber, 1974) ( $O(N \log N)$  complexity) or multigrid strategies (Brandt, 1977) ( $O(N)$  complexity). More details on the numerical strategy employed to compute the KR distance can be found in Métivier et al. (2016).

Regarding the implementation within the FWI framework, the modification of the misfit function from  $f_{L^2}$  to  $f_{KR}$  only impacts the way we design the adjoint source (Brossier et al., 2010; Luo and Sava, 2011). This adjoint source is given by

$$\partial_{d_{pred}} \|d_{pred}[m] - d_{obs}\|_{KR} = \overline{\varphi}(x, t) \quad (11)$$

where  $\overline{\varphi}(x, t)$  is the solution of the maximization problem (8). As a consequence, within the minimization loop for the solution of the FWI problem (3), the solution of the problem (8) is computed only once per iteration, for evaluating both the misfit function and the adjoint source.

## CASE STUDIES

For the numerical experiments, an acoustic time-domain finite-differences code 4th order (Marmousi) and 8th order (Chevron)

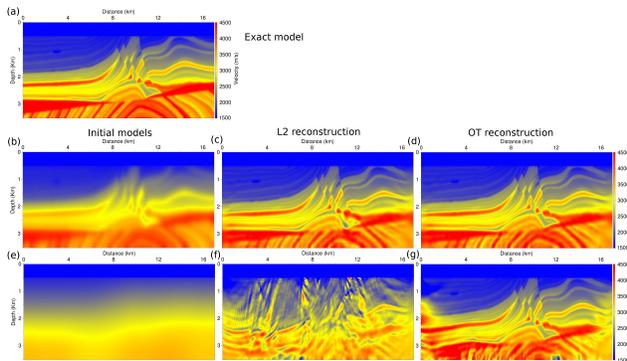


Figure 1: Results on the Marmousi model. Exact model (a), First initial model (b),  $L^2$  (c) and KR (d) reconstructions in the first initial model, second initial model (e),  $L^2$  (f) and KR (g) reconstructions in the second initial model.

in space, 2nd order in time is used. We start with an example of application on the Marmousi model. We consider a surface acquisition with 128 sources every 125 m and 168 receivers every 100 m. The source signature is a Ricker wavelet centered on 5 Hz from which the frequency content below 2.5 Hz has been removed using a Butterworth filter. We consider two initial models, obtained after smoothing the exact model with a Gaussian filter with a correlation length equal to 250 m and 2000 m respectively. The first initial model is close from the exact one, with smoother interfaces. The second initial model presents almost no lateral variations, and underestimates the velocity increase in depth. The results obtained by minimizing the  $L^2$  and KR distance starting from these two initial models are presented in Figure 1. Starting from the first initial model, minimizing the  $L^2$  and KR distance yields satisfactory reconstructions of the Marmousi model. Conversely, starting from the second model, the  $L^2$  estimation is far from the exact model, while the KR estimation is significantly closer, even if low and high velocity artifacts can be noted on the left part and in depth. This experiment indicates a better robustness of the KR distance with respect to cycle skipping.

A better insight on the KR distance is obtained by comparing the adjoint source in the second initial model corresponding to the  $L^2$  and KR misfit functions (Fig. 2). The KR distance yields smoother residuals, and re-weights the different seismic events such that a better balance between them is obtained. Lower amplitude reflections are amplified, and the resulting residuals share some similarities with a skeletonized version of the  $L^2$  residuals.

Next we consider the Chevron synthetic benchmark data-set issued in 2014. This data-set corresponds to a streamer acquisition, with a maximum of 8 km offset, with 321 receivers by sources equally spaced each 25 m. This data-set has been computed in the elastic approximation. We select 256 shots with a distance of 150 m between each sources. We use the initial model provided by Chevron: this is a 1D layered model incorporating the correct bathymetry. A low velocity layer between  $z = 2.3$  km and  $z = 3$  km can be observed. This velocity inversion and the relatively short available offsets (only 8 km) prevent diving waves from sampling the deepest part of the model. This makes the benchmark data challenging as only re-

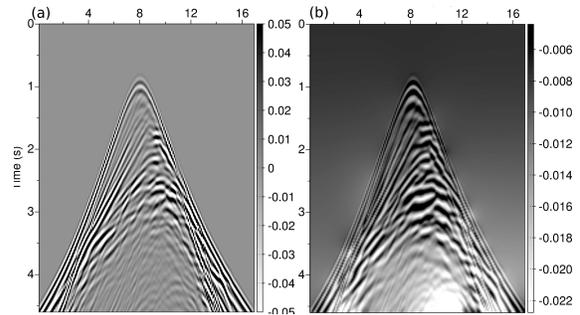


Figure 2:  $L^2$  (a) and KR (b) residuals in the second initial model for a shot gather with a source located at  $x = 8$  km.

flexion information is available for constraining the deep part of the model.

A simple workflow is implemented, relying only on a frequency continuation/multiscale strategy (Bunks et al., 1995): Butterworth low-pass and high-pass filters are applied to the selected shot-gathers to generate an ensemble of 15 data-sets with an increasing bandwidth from 2 – 4 Hz to 2 – 25 Hz. For each data-set, a wavelet estimation is performed using the initial velocity model, following the frequency-domain strategy introduced by Pratt (1999). The velocity model is discretized on a 2D mesh with discretization steps from  $h = 37.5$  m to  $h = 12.5$  m.

The results obtained up to 10 Hz using the KR distance and the  $L^2$  distance are presented in Figure 3. Following this very simple workflow the  $L^2$  distance clearly converges towards a local minimum. The velocity estimation at 10 Hz loses the tilted layered structure which was appearing in the  $L^2$  estimation at 4 Hz. Conversely, the KR estimation yields a coherent velocity estimation, which appears as a low resolution estimation of the final estimated model.

This final estimated model, obtained at 25 Hz using the KR distance, is presented in Figure 4. Three shallow low velocity anomalies are recovered at 500 m depth, as well as a small scale low velocity anomaly at  $x = 14.75$  km and  $z = 1$  km. The layered structure of the initial model is tilted in the final estimation. From left to right, the upper (faster) layers bend downward, while the low velocity layer at depth  $z = 2.5$  km bends upward. Three high velocity anomalies are recovered on top of the layer above the low velocity layer, at 1.8 km depth. The deeper part of the model ( $z > 3$  km) is less well reconstructed, as it could be expected from the lack of illumination of this zone. However, a curved interface is recovered at a depth between 4.5 and 5 km, as well as a flat reflector at the bottom of the model ( $z = 5.8$  km). This estimations displays strong similarities with existing results on this data-set, obtained either through a deconvolution misfit approach (Warner and Guasch, 2015) or with a  $L^2$  misfit starting from a high-resolution tomography velocity model (Qin et al., 2015).

To assess the quality of this estimation, a comparison of a synthetic and observed shot gathers at 25 Hz is presented in Figure 5. Mirroring the synthetic data allows to compare near offset reflections as well as far offset refracted data. The synthetic data is clearly in phase, and almost all the reflections are predicted. The amplitude of the main events is also cor-

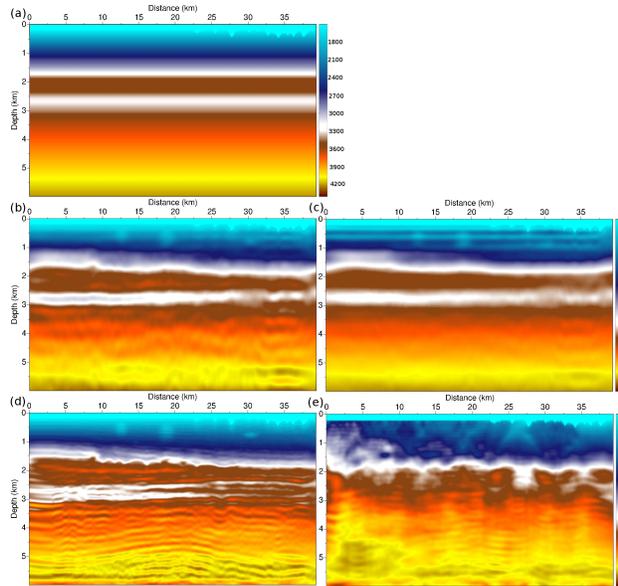


Figure 3: Initial 1D model (a). KR (b) and  $L^2$  (c) estimations at 4 Hz. KR (d) and  $L^2$  estimations at 10 Hz.

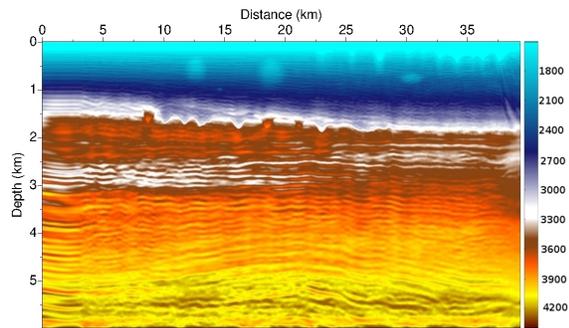


Figure 4: KR estimation obtained at 25 Hz.

rectly predicted, in the limit of the accuracy allowed by using an acoustic modeling engine to invert an elastic data-set. A slight time shift can be noted for the latest arrivals, which may be due to the accumulation of errors with propagating time, or an increasing kinematic inconsistency with large scattering angles, due to the use of an acoustic approximation to interpret elastic data.

The extra computational cost related to the computation of KR distance is presented in Table 1. In the worst case, (inversion of low frequency data-sets) the computational of the KR distance amounts to 49% of the total computation required for the computation of the gradient. This is equivalent to doubling the computation time compared to using the  $L^2$  distance. For higher frequencies, the mesh refinement is responsible for a faster increase of the computational complexity of the wavefield than the complexity of the computation of the KR distance, and the additional cost becomes lighter (only 8% of the computation of the gradient for the finest mesh). This is understandable as the computational complexity for wave modeling is in  $O(N_x \times N_z \times N_t)$  while for the proposed strategy for the KR norm it is in  $O(N_t \times N_r)$ , where the number of receivers  $N_r$  remains constant as the mesh is refined.

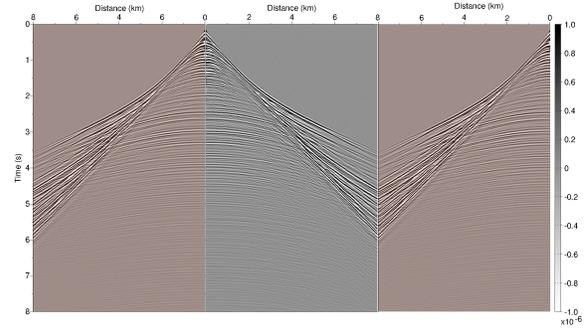


Figure 5: Comparison of the synthetic shot gather (orange background) with the benchmark shot gather at 25 Hz. The synthetic shot gather has been mirrored to ease the comparison at short and large offsets.

Mesh	$N_x \times N_z$	$N_t$	Gradient	KR norm- %
37.5 m	$160 \times 1310$	2001	86 s	42 s - 49%
25 m	$240 \times 1965$	2667	209 s	56 s - 26%
12.5 m	$480 \times 3930$	8001	2120 s	171 s - 8%

Table 1: Comparison between the total computational time for one gradient and the computational time for the KR norm within the gradient computation. The percentage of time dedicated to the computation of the KR norm decreases from 49% to 8% as the mesh is refined.

## CONCLUSION

The Kantorovich-Rubinstein distance appears as an interesting tool for mitigating cycle skipping in the framework of FWI. Using this method it is possible to relax the requirement either on the accuracy of the initial model (kinematic compatibility), and/or on the pre-processing of the data when using a  $L^2$  distance. Indeed, for the Chevron benchmark data-set, better results can be obtained with the  $L^2$  distance but require heavier pre-processing (for instance time/offset windowing strategies) and quality control steps to feed the inversion in an appropriate hierarchical order. In comparison, only a Bunks frequency continuation strategy is used here, and applying the KR distance seems to yield results comparable with state-of-the-art FWI technologies either based on the design of very accurate initial model through sophisticated high resolution tomography tools (Qin et al., 2015), or the use of a deconvolution based approach where the misfit function is defined as the penalization of energy not focused at zero time lag (Warner and Guasch, 2015). Further applications to real data (2D Nankai data (Operto et al., 2006), 3D Valhall data (Operto et al., 2015)) should give a better insight on the potentialities of the optimal transport strategy for FWI.

## ACKNOWLEDGEMENTS

This study was partially funded by the SEISCOPE II consortium, sponsored by BP, CGG, CHEVRON, EXXON-MOBIL, JGI, PETROBRAS, SAUDI ARAMCO, SCHLUMBERGER, SHELL, SINOPEC, STATOIL, TOTAL and WOODSIDE. See <https://seiscope2.osug.fr>. 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.

## EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2016 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

## REFERENCES

- Brandt, A., 1977, Multi-level adaptive solutions to boundary-value problems: *Mathematics of Computation*, **31**, 333–390, <http://dx.doi.org/10.1090/S0025-5718-1977-0431719-X>.
- Brossier, R., S. Operto, and J. Virieux, 2010, Which data residual norm for robust elastic frequency-domain full waveform inversion?: *Geophysics*, **75**, no. 3, R37–R46, <http://dx.doi.org/10.1190/1.3379323>.
- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**, 1457–1473, <http://dx.doi.org/10.1190/1.1443880>.
- Combettes, P. L., and J.-C. Pesquet, 2011, Proximal splitting methods in signal processing, in H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, eds., *Fixed-point algorithms for inverse problems in science and engineering*, Springer optimization and its applications: Springer New York 49, 185–212.
- Engquist, B., and B. D. Froese, 2014, Application of the wasserstein metric to seismic signals: *Communications in Mathematical Sciences*, **12**, 979–988, <http://dx.doi.org/10.4310/CMS.2014.v12.n5.a7>.
- Jannane, M., W. Beydoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M. Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, D. Trezeguet, and M. Xie, 1989, Wavelengths of earth structures that can be resolved from seismic reflection data: *Geophysics*, **54**, 906–910, <http://dx.doi.org/10.1190/1.1442719>.
- Kantorovich, L., 1942, On the transfer of masses: *Doklady Akademii Nauk USSR*, **37**, 7–8.
- Lailly, P., 1983, The seismic problem as a sequence of before-stack migrations, in J. Bednar, ed., *Conference on inverse scattering: Theory and applications*: SIAM.
- Lellmann, J., D. Lorenz, C. Schönlieb, and T. Valkonen, 2014, Imaging with Kantorovich-rubinstein discrepancy: *SIAM Journal on Imaging Sciences*, **7**, 2833–2859, <http://dx.doi.org/10.1137/140975528>.
- Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792, <http://dx.doi.org/10.1190/1.3627773>.
- Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: *Geophysics*, **56**, 645–653, <http://dx.doi.org/10.1190/1.1443081>.
- Ma, Y., and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and full waveform inversion: *Geophysics*, **78**, no. 6, R223–R233, <http://dx.doi.org/10.1190/geo2013-0004.1>.
- Métivier, L., R. Brossier, Q. Méridot, E. Oudet, and J. Virieux, 2016, Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion: *Geophysical Journal International*, **205**, 345–377, <http://dx.doi.org/10.1093/gji/ggw014>.
- Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and J. Virieux, 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**, 1362–1391, <http://dx.doi.org/10.1093/gji/ggv226>.
- Operto, S., J. Virieux, J. X. Dessa, and G. Pascal, 2006, Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: Application to the eastern

- Nankai trough: *Journal of Geophysical Research*, **111**, B09306, <http://dx.doi.org/10.1029/2005JB003835>.
- Plessix, R. E., and C. Perkins, 2010, Full waveform inversion of a deep water ocean bottom seismometer dataset: *First Break*, **28**, 71–78, <http://dx.doi.org/10.3997/1365-2397.2010013>.
- 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, <http://dx.doi.org/10.1190/1.1444597>.
- Qin, B., T. Allemand, and G. Lambaré, 2015, Full waveform inversion using preserved amplitude reverse time migration: 85th Annual International Meeting, SEG, Expanded Abstracts, 1252–1257, <http://dx.doi.org/10.1190/segam2015-5865645.1>.
- Swarztrauber, P. N., 1974, A direct method for the discrete solution of separable elliptic equations: *SIAM Journal on Numerical Analysis*, **11**, 1136–1150, <http://dx.doi.org/10.1137/0711086>.
- Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical Prospecting*, **56**, 765–790, <http://dx.doi.org/10.1111/j.1365-2478.2008.00698.x>.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266, <http://dx.doi.org/10.1190/1.1441754>.
- Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26, <http://dx.doi.org/10.1190/1.3238367>.
- Warner, M., and L. Guasch, 2014, Adaptive waveform inversion — FWI without cycle skipping — Theory: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, <http://dx.doi.org/10.3997/2214-4609.20141092>.
- Warner, M., and L. Guasch, 2015, Robust adaptive waveform inversion: 85th Annual International Meeting, SEG, Expanded Abstracts, 1059–1063, <http://dx.doi.org/10.1190/segam2015-5853026.1>.
- Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform inversion: *Geophysics*, **78**, no. 2, R59–R80, <http://dx.doi.org/10.1190/geo2012-0338.1>.