Mitigating cycle skipping in FWI through preconditioned multi-dimensional optimal transport

James McLeman*, Tim Burgess, Tom Rayment, and Victor Chua, DUG Technology

Summary

Many implementations of full waveform inversion (FWI) use an L2 norm to quantify differences between the modelled and observed data. Unfortunately, this approach is prone to cycle skipping when the supplied velocity is too far from the truth. Starting FWI at lower frequencies widens the desired basin of attraction yielding convergence towards a more correct solution. This, however, relies on the availability of good low frequency signal in the recorded seismic data. An alternative approach is to consider a more convex objective function, such as the Kantorovich-Rubinstein (KR) distance found in optimal transport problems. Determining the KR distance in multi-dimensional (multi-trace) optimal transport is tractable, but computationally expensive for many iterations and so we demonstrate an efficient implementation using a preconditioned LSQR approach together with the proximal splitting simultaneous-direction method of multipliers (SDMM) optimization technique. We compare the KR and L2 norms on data from the North-West Shelf of Australia.

Introduction

FWI seeks to minimize differences between the observed data and data modelled using the wave equation with initial estimates of subsurface models (Tarantola, 1984; Plessix, 2009). In doing so, the model(s) (e.g., P-velocity) are iteratively updated to explain these differences, increasing model resolution in the process. The mismatch between the modelled and observed data is typically quantified using an L2 norm of their difference. Minimizing the L2 norm is often achieved through local optimization quasi-Newton methods (Nocedal and Wright, 2006; McLeman et al., 2021), which use estimates of the inverse Hessian operator.

The L2 norm is computed sample by sample and hence is not well suited to measure time-shifts between the modelled and observed data due to their oscillatory nature. It is highly non-convex and so for large time-shifts, more than half a period, local optimization techniques will tend to converge to a local minimum, yielding an incorrect solution (Virieux and Operto, 2009). This is the notorious "cycle skipping" problem. The L2 norm is also highly sensitive to amplitude differences, which can occur from acoustic modelling inaccurately capturing the elastic-like dynamics of the subsurface. A common strategy to mitigate L2-norm issues is to initially consider only the amplitude-normalized diving waves and start FWI at low frequencies, thus widening the basin of attraction of the global minimum. With this first updated model, the inversion frequency can be increased and FWI run again (Operto et al., 2004). This workflow repeats until a desired frequency is reached. The process is time-consuming and relies on the availability of good low-frequency signal in the observed data.

Several data-domain methods with more convex objective functions were developed as an alternative. These prioritize travel-time differences between the modelled and observed data, rather than all aspects of the residual. The wave-equation travel-time inversion approach (Luo and Schuster, 1991) estimates time-shifts from the cross-correlation of the modelled and observed data but resulting model updates were typically lower resolution. Subsequent developments utilized an objective function based on a deconvolution approach (Luo and Sava, 2011; Warner and Guasch, 2016) which alleviated the assumption inherent in the cross-correlation approach that the seismic source was impulsive.

Mitigating cycle skipping in FWI with optimal transport (OT) was first discussed by Engquist and Froese (2014). Métivier et al. (2016a, 2016b, 2016c) further developed this using the dual formulation of the 1-Wasserstein distance, where the distance metric is given by the Kantorovich-Rubinstein (KR) norm (Villani, 2008). Unlike using a 2-Wasserstein distance (Wang & Wang, 2019), the KR norm is computationally tractable when considering data-domain multi-dimensional correlations in large-scale applications. The KR norm also does not require any data transformations to, for example, ensure "mass" positivity and conservation. Compared to the L2 norm, the KR norm has demonstrated improved convexity. Recent applications of the KR norm in FWI have shown success at an industrial scale (Messud and Sedova, 2019; Messud et al., 2021).

In this paper, we further demonstrate the application of multi-dimensional optimal transport in FWI using the KR norm. Its traditional computation with SDMM has been accelerated using the LSQR algorithm (Paige and Saunders, 1982a, 1982b) combined with a novel preconditioner. The results of this are shown on marine data from the North West Shelf of Australia and compared with a standard L2 norm.

Method

We begin with the 1-Wasserstein distance recast as a dual problem with a bounding constraint, yielding the KR norm (Messud and Sedova, 2019),

$$
\begin{aligned}
W^1(d_{mod}, d_{obs}) = \\
\max_{\varphi} \int_t \int_{x_{il}} \varphi(x_{il}, t)\big(d_{mod}(x_{il}, t) \\
- d_{obs}(x_{il}, t)\big) dx_{il} dt.
\end{aligned}
\tag{1}
$$

Multi-dimensional optimal transport in FWI

Subject to the constraints of,

$$|\varphi(x_{il}, t) - \varphi(x'_{il}, t')| \leq \gamma(x_{il}, t, x'_{il}, t') \quad (2)$$
$$|\varphi(x_{il}, t)| \leq \lambda \quad (3)$$

In equation (1), d_{mod} is the data synthesized through wave equation modelling, d_{obs} is the recorded data, x_{il} is a position along an inline (or along a cable in the marine towed streamer case) and t is time. Equation (2) is the 1-Lipschitz constraint which limits changes in $\varphi(x_{il}, t)$ to vary slower than $\gamma(x_{il}, t, x_{il}', t')$. Equation (3) represents a maximum bound on the amplitude of $\varphi(x_{il}, t)$, this constraint is essential for FWI applications since, in general, mass is not conserved and in such cases equation (1) will diverge. Equations (1)-(3) are a linear programming problem, but equation (2) represents a very large number of constraints. This can be avoided by choosing the metric $\gamma(x_{il}, t, x_{il}', t')$ to take the form of local constraints (Métivier et al., 2016b),

$$\gamma(x_{il}, t, x'_{il}, t') = \frac{1}{\sigma_{il}}(v|t^{k+1} - t^k| + |x^{p+1}_{il} - x^p_{il}|) \quad (4)$$

Where v is the velocity of the direction with the highest inter-trace correlations, σ_{il} is the variance along the inline, and k and p are sample indices in time and space respectively.

This linear programming problem can be discretized and iteratively solved using the SDMM optimization technique (Métivier et al, 2016b; Combettes and Pesquet, 2011). An expensive part of this optimizer involves solving a matrix inverse problem to determine the updated values of $\varphi(x_{il}, t)$ from the linear constraints. This is described by the solution of the normal equations,

$$\varphi^n = (I + L^T L)^{-1}[(y^n_1 - z^n_1) + L^T(y^n_2 - z^n_2)] \quad (5)$$

Where I is the identity matrix, L is the rectangular matrix of linear constraints, y and z are vectors generated during SDMM and n is the SDMM iteration number. The matrix $(I + L^T L)^{-1}$ is a large sparse matrix, too large to fit in memory for realistic shot gather sizes and therefore spectral decomposition methods are appealing. These, however, can still be computationally expensive in comparison to solving equation (5) in a preconditioned iterative fashion. We therefore recast equation (5) into the following form,

$$J = \min_x ||(I + L)Px^n - (y^n - z^n)||_2 \quad (6)$$

Where P is a linear preconditioning operator and $\varphi^n = Px^n$. We can construct P to dramatically reduce the number of iterations required to solve equation (6). The action of $(I + L)$ is known *a priori*, so an estimate of its point spread function (PSF) can be built by convolving $(I + L^T L)$ with a unit spike. The damped inverse and square root of the PSF amplitude spectrum gives our preconditioner P. The advantage of this is that P represents a filter, and its application is efficient since it is simply a convolution with x^n. We solve equation (6) using a preconditioned LSQR optimization scheme. P can be computed once, upfront, and then re-used for every LSQR, SDMM and FWI iteration.

A further efficiency can be achieved by recognizing that the adjoint source is given by,

$$\frac{\partial W^1(d_{mod}, d_{obs})}{\partial d_{mod}} = \varphi_{max}(x_{il}, t)$$
$$+ \int \int \frac{\partial \varphi_{max}(x'_{il}, t')}{\partial d_{mod}(x_{il}, t)}(d_{mod}(x'_{il}, t') - d_{obs}(x'_{il}, t'))dx'_{il}dt' \quad (7)$$

The second term in equation (7) can be considered negligible (Messud et al., 2021). The adjoint source is therefore just the $\varphi_{max}(x_{il}, t)$ already computed in equation (1) when determining the KR objective function value.

Results

Figure 1 shows a 6 Hz FWI synthetic experiment which compares the L2 and KR objective functions when an initial velocity model is perturbed via scaling by a constant ratio.
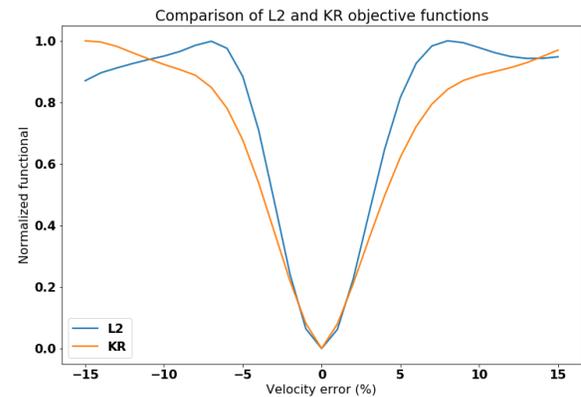
Figure 1: A comparison between the L2 objective function (blue) and the KR objective function (red).

The L2 objective function is non-convex, demonstrating multiple minima. In this scenario, a starting velocity error of +/-8% or more would yield convergence to an incorrect solution, as the downhill direction moves away from the global minimum. The inversion has cycle skipped. The KR objective function, however, is convex and even starting with a velocity error of +/-15%, FWI will iterate towards the correct solution. The widening of the global minimum basin

Multi-dimensional optimal transport in FWI

of attraction highlights the potential of this technique to make FWI more robust against cycle skipping.

We tested the KR objective function on a triple-source, 12-streamer dataset from the Australian North West Shelf. The initial model was generated from a single well check-shot survey which was then smoothed and extrapolated following the water-bottom horizon. FWI was then started at a frequency of 6 Hz, where only the diving waves were used to update the velocity model in a single parameter inversion. The water-bottom in this test area varied from 130 m – 415 m depth, with diving waves reaching a maximum penetration of 2 km. Starting FWI at this frequency with this initial model was known to cause cycle skipping in regions of the model when using an L2 objective function, making this a suitable test for the KR approach. This also determines the possibility of reducing time spent building up the low frequencies in conventional FWI workflows or having to start with a velocity derived from an RMO tomography update. For all results shown, 10 FWI iterations were used. In the KR FWI case, 40 SDMM iterations were used to compute the KR norm, each using approximately four preconditioned LSQR iterations.

The initial model, L2-updated model and KR-updated model overlain on their respective Kirchhoff prestack depth migration (preSDM) stacks are shown in Figure 2. L2 FWI has demonstrated some improvement in the shallow-water region but has also introduced some local undulations which are incorrect. The deeper-water region shows a degradation from the initial model, with non-geophysical velocity trends. This is not unexpected, as the well check-shot survey from which the initial model was derived was located in the shallow-water region ~1.2 km out of the displayed plane of section. This results in the deeper-water region having an initial model which is relatively further from the truth than the shallower water-bottom section. The KR FWI result demonstrates a dramatic improvement in imaging throughout the section, particularly in the deep-water region. The improvement in objective function convexity brought by the KR norm has allowed FWI to lock onto the correct solution even when starting very far from the truth.

Figures 3a), b), & c) show image gathers through the deeper-water section migrated with the initial, L2 FWI and KR FWI velocity models respectively. Figures 3d), e), & f) show the same but through a shallower water-bottom section. As is hinted at by the stack results, the image gathers show considerable move-out through the deeper water-bottom region in both the initial model and the L2 FWI derived model. On the other hand, KR FWI has slowed the velocity down in this region considerably which was needed to bring the inverted velocity model closer to what is required to flatten the image gathers. The improvement in gather flatness in the shallower water-bottom region is largely

comparable, with local exceptions, between the L2 and KR FWI approaches. Both are considerable improvements over the initial model.
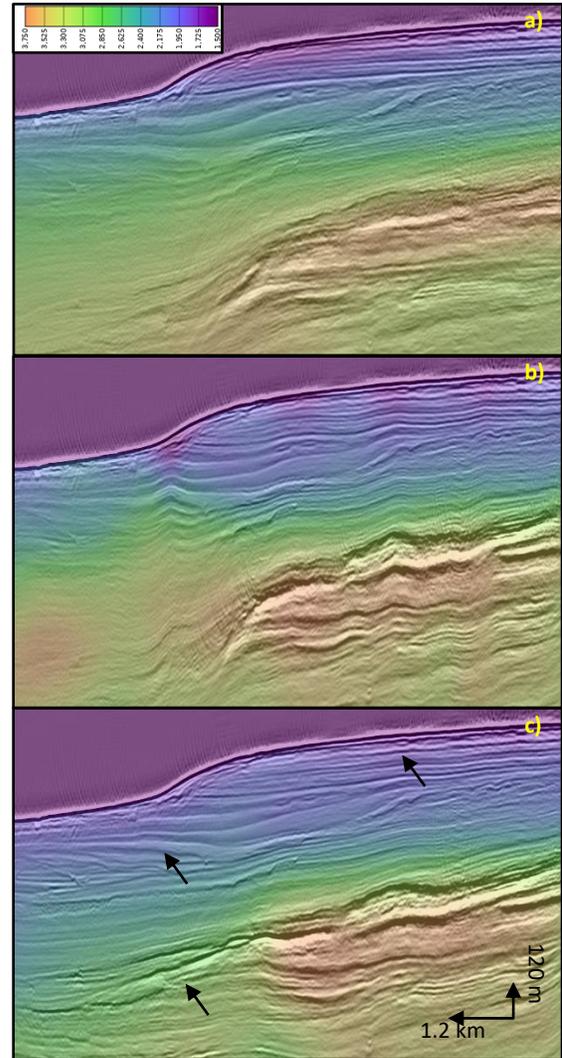
Figure 2: Kirchhoff preSDM stacks with their respective velocity models overlain where a) is the initial model, b) L2 FWI output, and c) KR FWI output.

Figure 4a) shows a comparison between the check-shot velocity (black), the initial velocity model (red) and the KR FWI result (cyan). The KR FWI updated velocity shows a good match with the check-shot profile, where additional resolution will be added upon increasing the FWI frequency. Figures 4b), c) & d) show the 1D cross-correlation between the observed data and data modelled using the initial velocity, L2 FWI velocity, and KR FWI velocity

Multi-dimensional optimal transport in FWI

respectively. Using the initial model, time-shifts between the modelled and observed data are apparent and increase with increasing offset. L2 FWI attempts to reduce this, but the obvious cycle skip has prevented it from obtaining the correct solution. The KR FWI demonstrates a flat, near-symmetric cross-correlation indicating that the kinematics of the modelled and observed data match well at all offsets.

Figure 4: a) Well profile comparing the check-shot (black), initial model (red) and OT FWI result (cyan). Cross-correlations between the modelled and observed data are shown for the b) initial model, c) L2 FWI, and d) KR FWI. The cross-correlogram values have been clipped to show only positive values

Conclusions

In this paper we have described the application of multi-dimensional optimal transport for mitigating cycle skipping in FWI. This discussion highlighted techniques that can be used to improve the efficiency of computing the KR norm, which forms a more convex objective function in comparison to the standard L2 norm approach. These novel techniques were applied to a real dataset from the Australian NW continental shelf. Despite starting from a smoothed 1D velocity profile derived from a single well check-shot survey, the KR FWI approach successfully converged towards a more correct solution. The resulting velocity model demonstrated a substantial improvement in imaging compared to the L2 FWI case. This highlights the possibility of starting FWI at higher inversion frequencies even with less accurate starting models. Thus, also reducing the number of FWI frequency steps that would traditionally be used in a standard FWI workflow and reducing the reliance on good recorded low frequencies.

Acknowledgements

Figure 3: Kirchhoff preSDM image gathers migrated with a) initial model, b) L2 FWI, and c) KR FWI through the deeper water-bottom region, d), e) and f) are the same but through the shallower water-bottom section.

REFERENCES

Combettes, P. L., and J.-C. Pesquet, 2011, Proximal splitting methods in signal processing, in Fixed-point algorithms for inverse problems in science and engineering: Springer, Optimization and Its Applications, **49**, 185–212.

Engquist, B., and B. Froese, 2014, Application of the Wasserstein metric to seismic signals: Communications in Mathematical Sciences, **12**, 979–988, doi: https://doi.org/10.4310/CMS.2014.v12.n5.a7.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792, doi: https://doi.org/10.1190/1.3627773.

Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653, doi: https://doi.org/10.1190/1.1443081.

McLeman, J., T. Burgess, M. Sinha, G. Hampson, and T. Thompson, 2021, Reflection FWI with an augmented wave equation and quasi-Newton adaptive gradient scheme: First International Meeting for Applied Geoscience & Energy, SEG/AAPG, Expanded Abstracts, 667–671, doi: https://doi.org/10.1190/segam2021-3582659.1.

Messud, J., R. Poncet, and G. Lambaré, 2021, Optimal transport in full-waveform inversion: Analysis and practice of the multidimensional Kantorovich-Rubinstein norm: Inverse Problems, **37**, 065012, doi: https://doi.org/10.1088/1361-6420/abfb4c.

Messud, J., and A. Sedova, 2019, Multidimensional optimal transport for 3D FWI: Demonstration on field data: 81st Conference and Exhibition, EAGE, Extended Abstracts, Tu R08 02, doi: https://doi.org/10.3997/2214-4609.201900869.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016a, Increasing the robustness and applicability of full waveform inversion: An optimal transport distance strategy: The Leading Edge, **35**, 1060–1067, doi: https://doi.org/10.1190/tle35121060.1.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016b, Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion: Geophysical Journal International, **205**, 345–377, doi: https://doi.org/10.1093/gji/ggw014.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016c, An optimal transport approach for seismic tomography: Application to 3D full waveform inversion: Inverse Problems, **32**, 115008, doi: https://doi.org/10.1088/0266-5611/32/11/115008.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization, 2nd ed.: Springer.

Operto, S., C. Ravaut, L. Improta, J. Virieux, A. Herrero, and P. Dell'Aversana, 2004, Quantitative imaging of complex structures from multifold wide aperture seismic data: A case study: Geophysical Prospecting, **52**, 625–651, doi: https://doi.org/10.1111/j.1365-2478.2004.00452.x.

Paige, C. C., and M. A. Saunders, 1982a, LSQR: An algorithm for sparse linear equations and sparse least squares: ACM Transactions on Mathematical Software, **8**, 43–71, doi: https://doi.org/10.1145/355984.355989.

Paige, C. C., and M. A. Saunders, 1982b, Algorithm 583: LSQR: Sparse linear equations and least squares problems: ACM Transactions on Mathematical Software, **8**, 195–209, doi: https://doi.org/10.1145/355993.356000.

Plessix, R.-E., 2009, Three-dimensional frequency-domain full-waveform inversion with an iterative solver: Geophysics, **74**, no. 6, WC149–WC157, doi: https://doi.org/10.1190/1.3211198.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, doi: https://doi.org/10.1190/1.1441754.

Villani, C., 2008, Optimal transport: Old and new: Springer Science & Business Media, 338.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, doi: https://doi.org/10.1190/1.3238367.

Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: Geophysics, **81**, no. 6, R429–R445, doi: https://doi.org/10.1190/geo2015-0387.1.