# Long-wavelength FWI updates beyond cycle skipping

*J. Ramos-Martínez\*, L. Qiu, J. Kirkebø and A.A. Valenciano, PGS, Y. Yang, UT-Austin*

## Summary

Full Waveform Inversion (FWI) success depends on producing seamless updates of the short- and long-wavelength features missing in the starting velocity model while avoiding cycle skipping. The use of cross-correlation gradients in FWI can lead to updates with the reflectivity imprint (high-wavenumbers) before the long wavelength updates have been constructed. In addition, the use of L2-norm to measure the data misfit is prone to cycle skipping. This may conduct to a local-minimum if the data lacks of low frequency information and/or the initial model is far from the true earth model. We offer a solution to these two FWI fundamental problems that combines a robust implementation of the velocity sensitivity kernel and the optimal transport norm to measure the data misfit. The new scheme can retrieve the long wavelength updates and reduce the cycle skipping problem. The velocity kernel eliminates the migration isochrones emphasizing the long-wavelength updates produced by the diving waves and the "rabbit ears" provided by reflections. The optimal transport norm accentuates those long-wavelength updates while minimizing the cycle skipping. We demonstrate the advantages of our implementation on synthetic and field data examples.

## Introduction

Classical FWI (Tarantola, 1984) can lead to velocity models with the reflectivity imprint. This is because the high-wavenumbers provided by the reflections often dominate the inversion over the low wavenumber updates (Mora, 1989). To minimize the problem, practitioners follow different data selection strategies to separate diving waves from reflections. However, the separation in the data space can be challenging. This, and the fact that for deep targets only reflections are available for the inversion, has motivated the development of FWI gradients that separate the wavenumber components in the velocity updates (e.g., Xu et al., 2013; Zhou et al., 2015; Ramos-Martínez et al., 2016).

Furthermore, FWI based on the L2-norm is an ill-posed and non-convex problem. The misfit function based on the L2 norm measures the difference between the recorded and modelled oscillatory signals in a point-by-point basis. This constrains FWI to use initial models that allow the simulation of waveforms within half of the period of the recorded waveforms. In case the initial model does not satisfy this condition, the inversion may suffer from cycle skipping and the solution will converge to the wrong

velocity model. In practice, this limitation can be overcome by applying a laborious data selection strategy. The events with the nearest offsets and the lowest possible frequencies are inverted first. In subsequent stages, increased offset ranges and broader frequency bandwidths are then considered. However, in many cases, the acquired seismic data do not have enough low frequencies to comply with the half of the period condition. Moreover, in complex geological settings such as those with the presence of salt, a small error in the location of the reflectors may lead to large kinematic errors. Thus, there is an incentive to change the metrics away from the L2 norm for quantifying the data misfit (e.g., Enquist et al., 2016; Métivier et al., 2016; Qiu et al., 2017).

Here, we adapt the velocity gradient for FWI to the optimal transport norm (W2) for measuring the data misfit. The numerical implementation introduces dynamic weights (Ramos–Martinez et al., 2016) in the velocity sensitivity kernel derived from impedance and velocity parameterization of the objective function. It effectively separates the migration isochrones produced by the specular reflectivity, from the components produced by the diving waves and the "rabbit ears". The new objective function was developed in the context of optimal transport theory. Our implementation (Qiu et al., 2017) uses an encoding scheme based on a logistic function that assures the mass conservation and positiveness required by the optimal transport theory.

## Theory

FWI is formulated as a nonlinear inverse problem matching modeled data to the recorded field data (Tarantola, 1984). Generally, a least-square objective function is used for measuring the data misfit between the modeled ($u$) and the recorded ($d$) data. Here we measure the data misfit using the quadratic form of the Wasserstein distance (W2 norm)

$$J = \sum_{shot}\sum_{rec} W_2^2(\widetilde{u},\widetilde{d}), \qquad (1)$$

where $\widetilde{u}(t)$ and $\widetilde{d}(t)$ are encoded versions of the modeled and field data. The definition of the quadratic Wasserstein distance and the resulting Frechet derivative to obtain the adjoint source are explained in Qiu et al. (2017).

We use the logistic function to encode both the field and modeled data. This function assures the mass balance and the positiveness conditions needed for the Wasserstein metrics. After the encoding, Cumulative Distribution

**Long-wavelength FWI updates beyond cycle skipping**

Functions (CDFs) are computed to obtain the adjoint source.

In order to produce long wavelength velocity updates we adapted the equations for the velocity gradient (Ramos-Martínez, 2016) to work with the W2 misfit function. The dynamic weight implementation of the velocity kernel was translated to the equivalent expressions such that the first-order time derivatives of the source and residual wavefields are computed before the adjoint source back propagation. The resulting velocity kernel has the form

$$G_V(\mathbf{x}) = \frac{1}{2A(\mathbf{x})}\left\{ \int_t \left[ W_1(\mathbf{x},t)\frac{1}{v^2(\mathbf{x})}S(\mathbf{x},t)R(\mathbf{x},T-t)\right]dt - \int_t \left[ W_2(\mathbf{x},t)\nabla S(\mathbf{x},t)\cdot\nabla\widetilde{R}(\mathbf{x},T-t)\right]dt \right\} \quad (2),$$

where $S(\mathbf{x},t)$ is the source wavefield; $R(\mathbf{x},T-t)$ is the wavefield computed from the adjoint-state equation, using the time reversal of a preconditioned version of the adjoint source. This preconditioning is done to compensate the DC bias introduced by the CDF's. $\widetilde{R}(\mathbf{x},T-t)$ results from applying the inverse operator of the preconditioner to the receiver wavefield. $A(\mathbf{x})$ is the illumination term. The dynamic weights $W_i(\mathbf{x},t)$ are designed to optimally suppress the unwanted specular reflectivity (migration isochrones).

Figure 1 shows the kernels for different combinations of the L2- and W2-norms, and the crosscorrelation and velocity gradients. These were computed for a source-receiver pair in a layer with an increasing velocity as a function of depth. Notice that the W2 velocity kernel accentuates the long-wavelength components corresponding to the diving waves and the "rabbit ears", compared to the L2 velocity kernel.

**Synthetic example**

We illustrate the advantages of using the W2 velocity kernel in FWI with a model consisting of four horizontal layers overlaying a half-space. The first layer is the water column up to 0.5 km depth. The other three layers have thicknesses of 1.5, 1.0 and 1.0 km, respectively, with different V(z) variations. During the inversion, we used both transmitted and reflected events. The frequency bandwidth of inversion has low- and high-cut frequencies between 3 to 13 Hz. The velocity difference between the true and the initial velocity model (Figure 2a) is large enough to produce cycle skipping in the reflected events corresponding to the second, third and fourth interfaces. This can be seen in the comparison of shot gathers computed for the true and initial models (Figure 3a). As observed, cycle skipping is even clearer in the long offset transmitted events.

We perform full-waveform inversion using the velocity gradient with the L2- and W2-norms. Velocity updates for the different options are shown in Figure 2. The FWI L2 velocity gradient (Figure 2b) provides the correct updates in the shallow part of the model, but the solution diverges in the deep. In contrast, the combination of the W2 norm and velocity gradient (Figure 2e) produces a good solution throughout the whole model.

For QC purposes, Figure 3b shows the comparison of waveforms computed with the true and inverted models using the W2 velocity gradient. Although the waveform fit is not perfect in amplitude, the FWI model provides a good phase agreement.
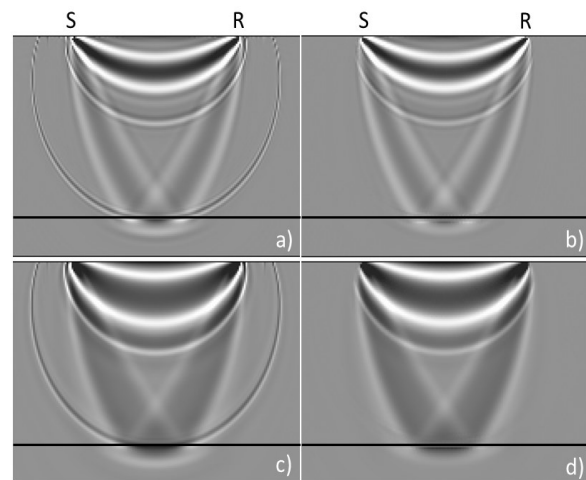


*Figure 1. Sensivity kernels of a source-receiver pair in a model with a V(z) layer overlaying a half-space for the a) L2-norm and crosscorrelation, b) L2-norm and velocity, c) W2-norm and crosscorrelation and d) W2-norm and velocity.*

**Field data example**

We compared the FWI results from the L2 velocity gradient and the W2 velocity gradient using field data. The data was acquired in deep-water Norwegian Sea. The acquisition comprises 16 dual-sensor streamers separated by 75 m, with a maximum inline offset of 8.1 km. The data has good signal-to-noise ratio at frequencies as low as 2 Hz. The FWI used a simple initial velocity model that fits water bottom reflections. The maximum high-pass frequency of the data used in the inversion is 9 Hz, in a window containing transmitted and reflected events.

In Figure 5a we show a comparison of the recorded and synthetic waveforms computed for the initial model. It clearly shows the cycle skipping at the waveforms

corresponding to the long-offset first arrivals. Figure 4b and 4c show the inverted models using the L2 velocity gradient and the W2 velocity gradient, respectively.

Figure 5b shows the comparison of the field and modeled traces with the inverted model using the L2 velocity gradient. Seemingly, cycle skipping is overcome in the far offsets. However, at intermediate offsets (~4 km) the model produces evident cycle skipping that is not present in the traces computed with the initial model. At the same time, the long offsets overlapped events in the field data are not reproduced by the model. In contrast, the inverted model with the W2 velocity gradient produces waveforms that overcome the cycle skipping at long offsets. This is achieved without hampering the waveform match at intermediate offsets. Moreover, the overlapped events observed in the recorded data, are reproduced.

## Conclusions

We combine a robust implementation of the velocity kernel and the optimal transport norm to overcome cycle skipping problem and retrieve long-wavelength velocity updates from transmitted and reflected events. The proposed solution expands the use of FWI for velocity model building because: it reduces the dependence on a good starting velocity model and produces long-wavelength updates from reflections. Thus, the requirements of long offsets and low-frequency input data are also reduced. We illustrate the advantages of our solution by using synthetic and field data from the Norwegian Sea.

## Acknowledgments

We thank PGS for permission to publish this material and Andrew Long for the critical reading of the paper. We appreciate helpful discussions with Nizar Chemingui and Faqi Liu, and the assistance of Sören Naumann, Øystein Korsmo and Mikhail Orlovich in the field data example.
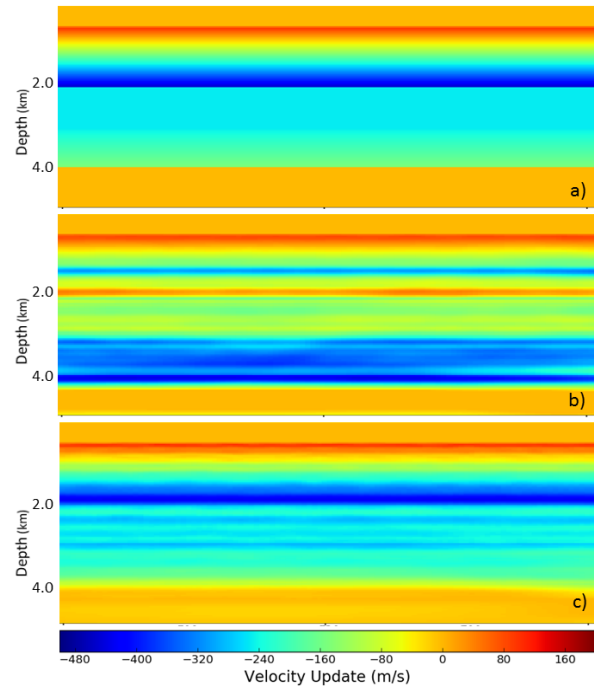


*Figure 2. a) Velocity difference between the true and initial models. FWI velocity updates using the b) L2-norm and velocity gradient, c) W2-norm and the velocity gradient.*
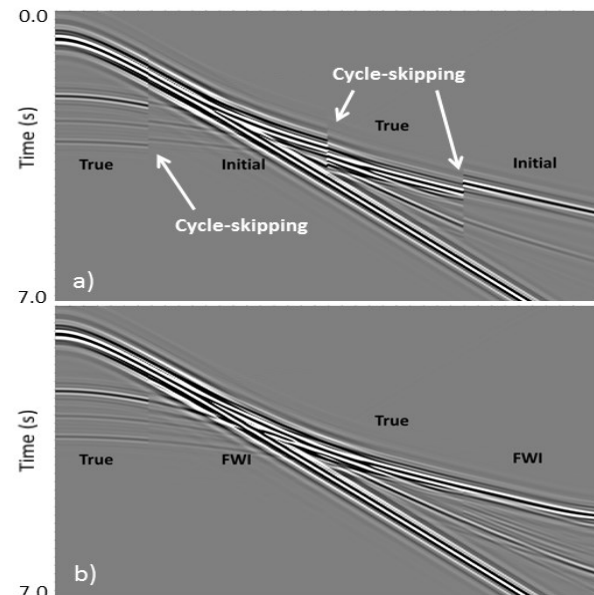


*Figure 3. Comparison of waveforms at different offset ranges computed from the a) true and initial model, and b) the true and the FWI model using the W2-norm and the velocity gradient. Maximum offset is 12 km.*

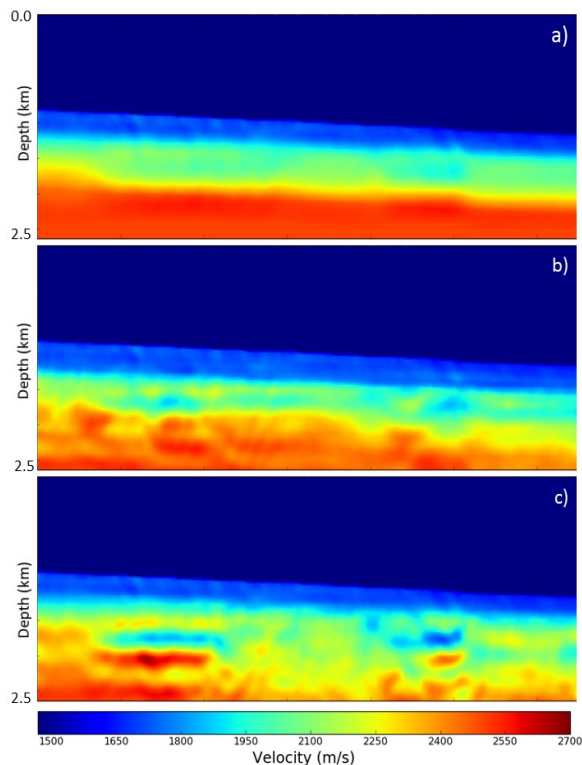**Long-wavelength FWI updates beyond cycle skipping**



*Figure 4. Vertical velocity profiles for the field data example: a) initial and inverted models using the b) L2-norm and velocity gradient, c) the W2-norm and velocity gradient. Horizontal distance is 32.5 km.*
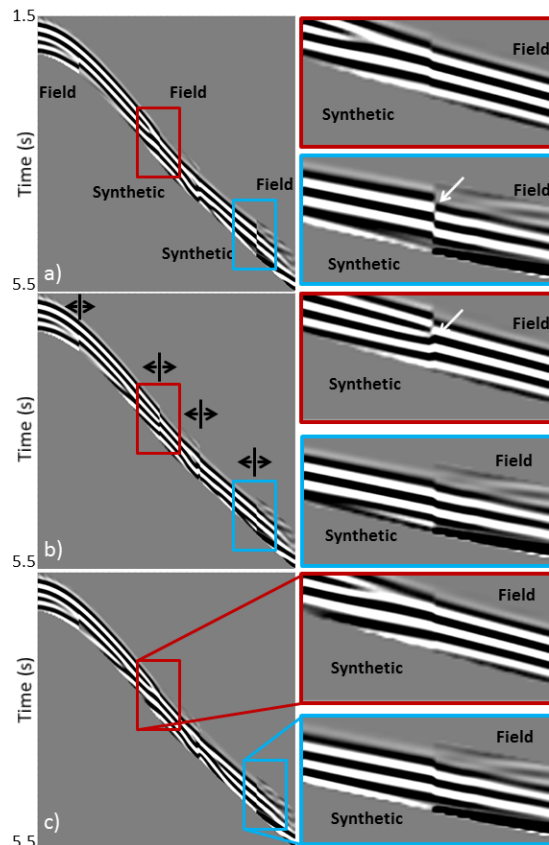


*Figure 5. Comparison of field and synthetic shot gathers at different offset ranges for one sample shot. Synthetic traces were computed with: (a) initial model, (b) inverted model using the velocity kernel and L2-norm, (c) inverted model using the velocity kernel and W2 norm. Two zoomed windows for intermediate (red) and far (blue) offsets are displayed at the right of each panel. The boundaries between the field and synthetic waveforms are marked with black arrows and the cycle skipping with white arrows.*

# REFERENCES

Enquist, B., B. D. Froese, and Y. Yang, 2016, Optimal transport for seismic full waveform inversion: Communications in Mathematical Sciences, **14**, 2309–2330, https://doi.org/10.4310/CMS.2016.v14.n8.a9.

Mora, P., 1989, Inversion = migration + tomography: Geophysics, **54**, 1575–1586, https://doi.org/10.1190/1.1442625.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Vireux, 2016, Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion, Geophysical Journal International, **205**, 345–377.

Qiu, L., J. Ramos-Martinez, A. A. Valenciano, Y. Yang, and B. Enquist, 2017, Full-waveform inversion with an exponentially encoded optimal-transport norm: 87th Annual International Meeting, SEG, Expanded Abstracts, 1286–1290, https://doi.org/10.1190/segam2017-17681930.1.

Ramos-Martinez, J., S. Crawley, Z. Zou, A. A. Valenciano, L. Qiu, and N. Chemingui, 2016, A robust gradient for long wavelength FWI updates: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, 547–551.

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

Xu, S., D. Wang, F. Chen, Y. Zhang, and G. Lambaré, 2013, Full waveform inversion for reflected seismic data: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts, https://doi.org/10.3997/2214-4609.20148725.

Zhou, W., R. Brossier, S. Operto, and J. Vireux, 2015, Full waveform inversion of diving waves for velocity model building with impedance inversion based on scale separation: Geophysical Journal International, **202**, 1535–1554, https://doi.org/10.1093/gji/ggv228.