Adaptive quadratic Wasserstein full-waveform inversion

Diancheng Wang and Ping Wang, CGG*

Summary

Full-waveform inversion (FWI) has increasingly become standard practice in the industry to resolve complex velocities. However, the current FWI research still exhibits a diverging scene, with various flavors of FWI targeting different aspects of the problem. Outstanding challenges currently faced by FWI include severe cycle-skipping and amplitude-mismatch situations, such as in salt updates, and limited acquisition cases where mainly reflection arrivals are present. Recent advances in FWI to update salt show that with the proper handling of cycle-skipping issues and amplitude effects, FWI can realize its potential of correcting large, complex velocity errors, given adequate data. In this work, we present an FWI scheme based on the quadratic Wasserstein metric, with adaptive normalization and integral wavefield. We show that this scheme has better convexity than traditional metrics, and therefore can mitigate cycle-skipping issues, while also being insensitive to amplitude effects. We demonstrate the effectiveness of our approach with a streamer data set in an area of complex salt geometry. In addition, we show that this approach can also work naturally on reflection data, without the extra procedure of scale separation by either decomposition or demigration. With these learnings, we believe that the fundamentals of FWI research are starting to converge.

Introduction

Full-waveform inversion (FWI), initially proposed more than three decades ago (Tarantola, 1984), aims to resolve complex subsurface velocities using the full waveform information of seismic data. By iteratively updating the subsurface model in a way that minimizes the difference between synthetic and recorded seismic data, FWI has the potential to produce accurate and high-resolution velocity models. Throughout the last decade, the FWI community has witnessed a surge in both research activities and successful applications of FWI on industry-scale projects. Thanks to these advances, FWI has now been established as the method of choice for challenging model building tasks.

However, today's FWI research still exhibits a diverging scene, with various flavors of FWI co-existing, each usually targeting specific aspects of the overall problem. This is understandable considering that FWI is a highly ill-posed and non-linear process, with various difficulties intertwined together. Here we would like to point out several challenges that we think are most fundamental to FWI. One is the well-known cycle-skipping issue, which often traps the FWI update in local minima. The other is amplitude effects, which is often due to noise and wave modes in real data that cannot be modeled by the FWI wave-propagation engine. In addition, limited acquisition with insufficient offset range and angle coverage also poses a great challenge to FWI in general, making it a highly under-determined problem.

Salt update, one of the most difficult situations for FWI, can be considered the "poster boy" for cycle-skipping and amplitude-mismatch issues. Salt misinterpretation often causes large velocity errors that, in turn, result in severe cycle-skipping. And due to the strong contrast of salt boundaries, the amplitudes of salt events are often difficult to model correctly. Conventional FWI based on the least-squares (L^2) metric is known to be prone to both issues, as it tries to measure the misfit based on a pointwise amplitude difference. Recently, different approaches that aim to mainly capture the traveltime difference for inversion have shown considerable success (Warner and Guasch, 2014; Jiao et al., 2015; Zhang et al., 2018). The logic behind these methods is often to use a cost function that downplays the amplitude effects and mitigates cycle-skipping issues via either better utilization of low frequency or improved convexity. Given adequate data such as ocean bottom node (OBN), which has good low frequency signal and long offset range, these methods can generalize their success.

Even with a proper traveltime-based cost function, FWI can still be difficult when the acquisition is limited. FWI usually works well with diving waves, which provide good illumination and angle coverage within their penetration depth. However, not all acquisitions have the luxury of long offsets that OBN acquisition does. Oftentimes for streamer acquisition, we need to rely on reflection energies for deep velocity updating. To properly utilize the reflection energy for FWI, a scale separation procedure by either decomposition (Gomes and Chazalnoel, 2017) or demigration (Wang et al., 2018) is usually employed to separate the tomographic term from the migration term of the gradient. The tomographic term of the gradient, which represents kinematic information, is relevant to velocity errors and therefore kept for velocity updating. Recent progress in reflection FWI (RFWI) has demonstrated its effectiveness in improving image focusing and gather flatness. Nonetheless, the problem remains under-determined and the so-called "depth-velocity ambiguity" persists.

Optimal transport was first introduced to the geophysics community by Engquist et al. (2016) and has attracted considerable attention for its potential in FWI applications. The measure used in optimal transport, the Wasserstein distance, aims to find a transport map that matches two mass densities in a cost-efficient way. The fact that the Wasserstein distance encodes both the traveltime and amplitude information makes it a desirable candidate for replacing the L^2 metric in FWI. As shown by Engquist et al.

10.1190/segam2019-3214402.1

(2016), the Wasserstein distance has superior mathematical properties, including good convexity and insensitivity to amplitude distortions, which are well-suited for tackling the fundamental challenges FWI faces. Currently, there are mainly two different formulations of the FWI problem with optimal transport, one based on the 2-Wasserstein (W^2) metric (Yang et al., 2018) and the other based on the 1-Wasserstein (W^1) metric approximated by the Kantorovich-Rubinstein norm (Métivier et al., 2016). For both formulations, numerical simulations have illustrated that optimal transport can effectively mitigate cycle-skipping issues. However, successful applications of Wasserstein metric-based FWI on real industry data is still limited (Poncet et al., 2018; Ramos-Martínez et al., 2018), especially for salt velocity updating.

Here, we propose an adaptive quadratic Wasserstein (W^2) metric to further advance the effectiveness of optimal transport FWI. To properly apply the W^2 metric on seismic data, both the synthetic and real data need to be normalized to satisfy the positivity and mass-conservation requirements. The normalization scheme plays a critical role for the success of the algorithm and is where our innovation lies. In the following sections we will describe the details of this method and demonstrate its effectiveness with synthetic and field data examples.

Method

First we lay out the ground theory of FWI based on the quadratic Wasserstein metric. In this study, we computed the W^2 distance trace-by-trace for simplicity (We note that all the derivation below can be readily extended to multi-dimensions). Taking the synthetic data f and real data g as two density distributions defined on the same domain, we can compute the W^2 distance between these two distributions:

$$W^2(f,g) = \int_0^T \left|t - G^{-1}(F(t))\right|^2 f(t)dt, \qquad (1)$$

where $F(t) = \int_0^t f(s)ds, G(t) = \int_0^t g(s)ds$. It can also be expressed in matrix form:

$$W^2(f,g) = (t - G^{-1}(F(t)))^T diag(f)(t - G^{-1}(F(t))). \qquad (2)$$

In order to apply the W^2 cost function within the FWI framework, we also need to compute its Fréchet derivative with respect to the synthetic data f, i.e., the adjoint source. The adjoint source can be computed according to:

$$\nabla W^2 = \left[-2diag\left(\frac{dG^{-1}(y)}{dy}|_{F(t)}\right)Udiag(f) + diag\left(t - G^{-1}(F(t))\right)\right](t - G^{-1}(F(t))), \qquad (3)$$

where $\left(\frac{dG^{-1}(y)}{dy}|_{F(t)}\right) = \frac{1}{g(G^{-1}(F(t)))}$ and U is the upper triangle matrix. Once we obtain the adjoint source, we can back-propagate it to compute the gradient according to the adjoint-state method.

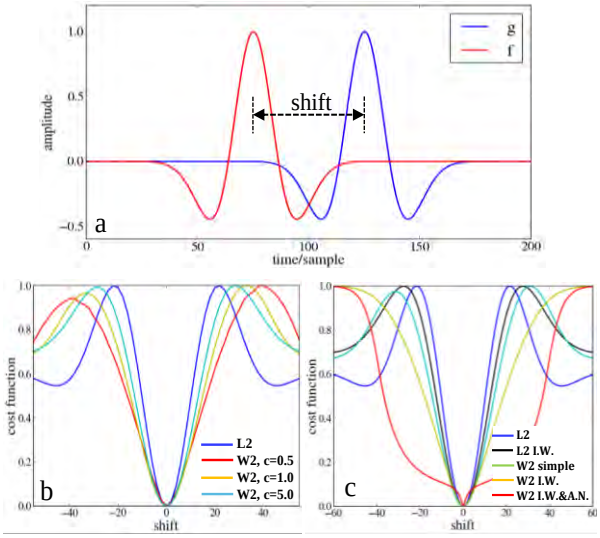

Figure 1: (a) synthetic signal f and observed signal g. Comparison of cost functions vs. time shift of L^2 metric with W^2 metric for (b) different normalization constant c and (c) integral wavefield and adaptive normalization.

For the W^2 metric FWI to work properly, the density distributions f and g need to satisfy two conditions: positivity and mass conservation. Both f and g have to be non-negative everywhere and their total masses must equal each other. Seismic data intrinsically has both positive and negative parts and there is no guarantee that the total masses of synthetic and real data will be equal. Therefore, normalization on data is necessary before applying the W^2 metric. Various normalization schemes have been discussed in literature (Qiu et al., 2017; Yang et al., 2018), each having its pros and cons. In this study we chose to use linear normalization. Experience from tests convinced us that linear normalization, although appearing to be the simplest, is most reliable when dealing with real data. The drawback from linear normalization is that it by nature weakens the convexity of the W^2 cost function. To mitigate this, we applied the following transformation to the seismic data before linear normalization:

$$\tilde{f}(t) = \int_0^t f(s)ds - \int_0^t g(s)ds, \quad \tilde{g}(t) = 0. \qquad (4)$$

We then applied linear normalization to the transformed traces $\tilde{f}$ and $\tilde{g}$:

$$\hat{f} \equiv \frac{\tilde{f} + c}{<\tilde{f} + c>}, \quad \hat{g} \equiv \frac{\tilde{g} + c}{<\tilde{g} + c>}, \qquad (5)$$

where c is a normalization constant that ensures $\tilde{f} + c > 0$ and $\tilde{g} + c > 0$, and $\langle * \rangle$ refers to the averaging operation. The normalized signals $\hat{f}$ and $\hat{g}$ were then plugged into Equations 1-3 for FWI velocity updating.

10.1190/segam2019-3214402.1

The transformation in Equation 4 contains two procedures: 1) integrating the wavefield (I.W.), which is intuitive since integration naturally boosts low frequency that enhances convexity, and (2) subtracting the integrated signals and matching the subtracted result to a trace with an average amplitude of zero, referred to as adaptive normalization (A.N.). The reasoning and effect of Equation 4 is illustrated in Figure 1. We used two identical Ricker wavelets separated by a certain time shift (Figure 1a) as the synthetic and real data traces. The W^2 cost function was plotted against the time shift. Figure 1b shows that W^2 with simple linear normalization loses global convexity, but is still better than L^2. In addition, the convexity of W^2 weakens as the normalization constant c increases. Therefore, we would like c to be as small as possible while still satisfying the positivity requirement. The adaptive normalization step in Equation 4 allows us to use a smaller c by taking a big chunk of energy out of the traces. The benefit from this becomes more clear when we consider that real seismic traces contain multiple events, with some significantly stronger than others. With adaptive normalization, we effectively remove the already matched events, which are usually strong, from the problem and gradually let the weaker events come into play. This allows a smaller c value and a boosted convexity. Figure 1c shows the overall benefit of Equation 4. We observed that integral wavefield improved the convexity for both L^2 (from blue to black) and W^2 (from green to yellow), and adaptive normalization was able to further improve the convexity for the W^2 metric (red).

In Figure 2 we show a synthetic example to validate our method. The simulation was performed on the Marmousi model (Figure 2a). A Ricker wavelet peaked at 4 Hz was used to generate the synthetic shots, with a fixed-spread acquisition geometry covering the model. The inversion was carried out at a single frequency of 6 Hz, with a 1D initial velocity model $v(z)$. Inversion results are shown in Figures 2b-d. We can see that the L^2 inversion is severely cycle-skipped. Simple W^2 inversion alleviated the cycle-skipping issue to some extent, while W^2 with integral wavefield and adaptive normalization could better recover the model, thanks to its superior convexity. In the following sections we will refer to W^2 inversion with adaptive normalization and integral wavefield as W2FWI.

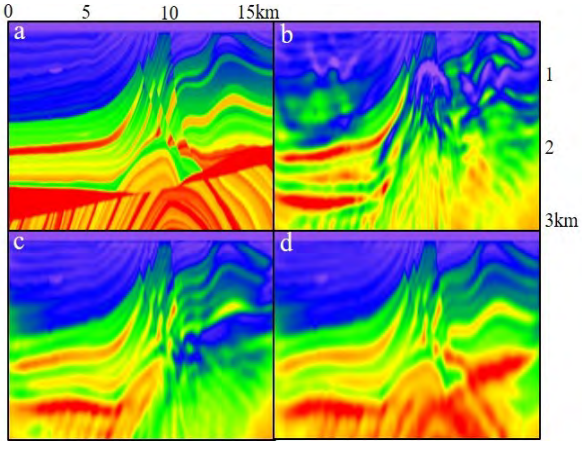

Figure 2: Synthetic tests on Marmousi model: (a) true model; (b) L^2 inversion; (c) simple W^2 inversion; and (d) W^2 inversion with adaptive normalization and integral wavefield.

Field data examples

W2FWI for salt update

The first field data example is from South Keathley Canyon in the Gulf of Mexico (GOM). The data was acquired with a staggered-vessel configuration, which provided full-azimuth coverage for offsets up to 10 km, and long offsets up to 18 km for certain azimuths. The area contains complex shallow salt, which is difficult for conventional L^2-based FWI. The initial model to W2FWI was a legacy model (Figures 3a-b), which came from the typical model building workflow of tomography and salt interpretation. The inversion was carried out from 3 to 7 Hz, and the results are shown in Figures 3c-d (direct output from W2FWI). We can see that W2FWI made significant corrections to the salt geometry, and as a result, the RTM image is much improved compared to the legacy image, especially for the subsalt events.

We note that similar results can also be achieved for this data set using other methods, e.g., the time-lag FWI (TLFWI) approach discussed by Zhang et al. (2018). At first thought, it is not immediately clear that these two drastically different methods should converge to similar results. Based on our current understanding, we believe that both W2FWI and TLFWI are able to better handle the two main issues of cycle-skipping and amplitude effects than conventional L^2-based FWI approaches. With these two issues properly addressed, the real power of FWI can then kick in and drive

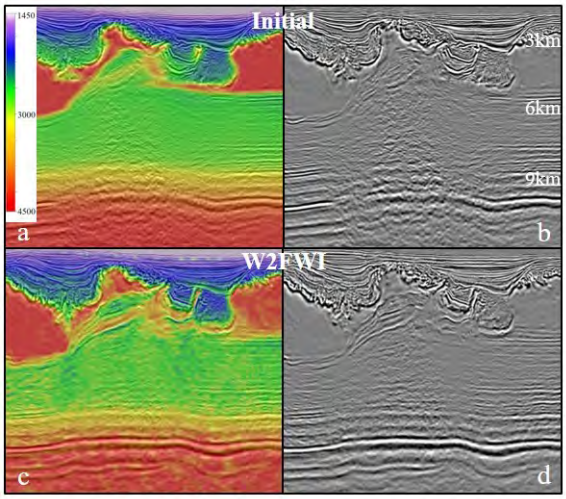

Figure 3: Field data from South Keathley Canyon: model and RTM image of (a)/(b) initial model and (c)/(d) W2FWI.

10.1190/segam2019-3214402.1

the model to a similar end point that best fits the data and improves the subsalt imaging.

W2FWI with reflection data

Next we look at a wide-azimuth (WAZ) data set from the GOM that has offset coverage up to about 8 km. With a limited offset range, diving waves cannot penetrate deep enough and we need to rely on reflections to update the deep velocity. It is well known that the gradient from reflection energy contains a strong migration term that obscures the tomographic term, which is much weaker in amplitude but more relevant for correcting kinematics. Typical RFWI solutions pick out the tomographic term for velocity update using a scale separation procedure by either decomposition or demigration. In this test, we decided to let W2FWI fully drive the inversion and kept the whole gradient as is. The results are shown in Figure 4. For comparison purposes, we also show the result from the CHF-RFWI solution, which uses the image-domain single-iteration least-squares reverse time migration (LSRTM) image obtained using curvelet-domain Hessian filter (CHF) (Wang et al., 2016) as the reflectivity (Gomes and Yang, 2018). We see that overall W2FWI provided a similar level of improvements to the image as CHF-RFWI. We do notice some minor differences in the deep section where W2FWI seems to provide smoother structures, which could be due to better angle resolution of W2FWI because it uses all of the data whereas RFWI only uses the reflection energy.

Figure 5 shows velocity updates from CHF-RFWI and W2FWI. At first look, the two seem to be very different. This is actually as expected since W2FWI retains the full gradient, so its update contains the strong high-wavenumber migration term unlike CHF-RFWI. When we look closer and through the migration term, we can see that the long wavelength update of W2FWI closely resembles that of CHF-RFWI, especially in the depth region between 4 and 9 km. It is this long wavelength component that made the major contribution to the image improvement.

The high-wavenumber migration term in the W2FWI update served two purposes. First, it acted as the reflectors necessary for generating reflections in the synthetic data. Second, it was a "garbage collector" that absorbed any misfit between real and synthetic data that could not be attributed to velocity errors. The latter made us suspect that caution needs to be exercised when trying to directly interpret high-resolution FWI models (Lu, 2016; Shen et al., 2018; Wang et al., 2019).

Conclusions

We presented a new FWI method based on the W^2 metric with adaptive normalization and integral wavefield. The effectiveness of this method was demonstrated with synthetic and field data examples. With the ability to better handle cycle-skipping and amplitude effects, this method can tackle challenging FWI problems, such as resolving complex salt geometries. In addition, the new method can naturally take advantage of the tomographic term from reflection data. Combining our learnings with knowledge from other successful FWI methods, we conclude that as long as the main issues with FWI are properly addressed, different methods can all reach a similar point of success. Therefore, we believe that after years of diverging efforts, the fundamentals of FWI research are starting to converge.

Acknowledgments

We thank CGG Multi-client & New Ventures for permission to publish this work. We also thank Xuhui Luo and Jun Zhou for help with the field data tests. We are grateful to Qing Xu, Zhigang Zhang, and Adriano Gomes for fruitful discussions.

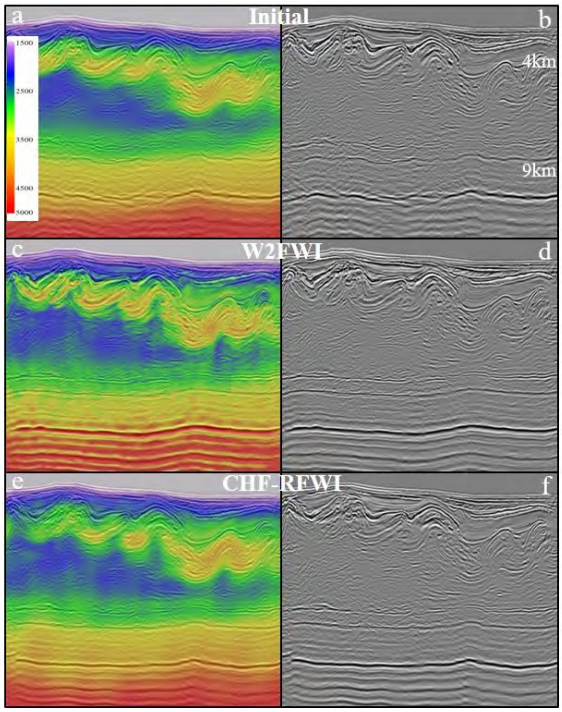

Figure 4: Field data from GOM: model and RTM image of **(**a)/(b) initial model, (c)/(d) W2FWI, and (e)/(f) RFWI with reflectivity from curvelet-domain single-iteration LSRTM image.

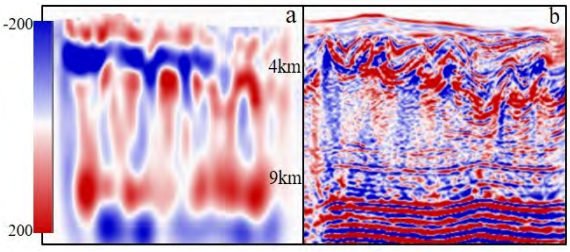

Figure 5: Field data from GOM: velocity update from (a) CHF-RFWI and (b) W2FWI.

10.1190/segam2019-3214402.1

REFERENCES

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

Gomes, A., and N. Chazalnoel, 2017, Extending the reach of full-waveform inversion with reflection data: Potential and challenges: 87th Annual International Meeting, SEG, Expanded Abstracts, 1454–1459, doi: https://doi.org/10.1190/segam2017-17731403.1.

Gomes, A., and Z. Yang, 2018, Improving reflection FWI reflectivity using LSRTM in the curvelet domain: 88th Annual International Meeting, SEG, Expanded Abstracts, 1248–1252, doi: https://doi.org/10.1190/segam2018-2996291.1.

Jiao, K., D. Sun, X. Cheng, and D. Vigh, 2015, Adjustive full waveform inversion: 85th Annual International Meeting, SEG, Expanded Abstracts, 1091–1095, doi: https://doi.org/10.1190/segam2015-5901541.1.

Lu, R., 2016, Revealing overburden and reservoir complexity with high-resolution FWI: 86th Annual International Meeting, SEG, Expanded Abstracts, 1242–1246, doi: https://doi.org/10.1190/segam2016-13872562.1.

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, doi: https://doi.org/10.1093/gji/ggw014.

Poncet, R., J. Messud, M. Bader, G. Lambaré, G. Viguier, and C. Hidalgo, 2018, FWI with optimal transport: A 3D implementation and an application on a field dataset: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts, We A12 02, doi: https://doi.org/10.3997/2214-4609.201801029.

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

Ramos-Martínez, J., L. Qiu, J. Kirkebø, A. A. Valenciano, and Y. Yang, 2018, Long-wavelength FWI updates beyond cycle skipping: 88th Annual International Meeting, SEG, Expanded Abstracts, 1168–1172.

Shen, X., L. Jiang, J. Dellinger, A. Brenders, C. Kumar, M. James, J. Etgen, D. Meaux, R. Walters, and N. Abdullayev, 2018, High-resolution full-waveform inversion for structural imaging in exploration: 87th Annual International Meeting, SEG, Expanded Abstracts, 1098–1102, doi: https://doi.org/10.1190/segam2018-2997202.1.

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

Wang, P., A. Gomes, Z. Zhang, and M. Wang, 2016, Least-squares RTM: Reality and possibilities for subsalt imaging: 86th Annual International Meeting, SEG, Expanded Abstracts, 4204–4209, doi: https://doi.org/10.1190/segam2016-13867926.1.

Wang, P., Z. Zhang, J. Mei, F. Lin, and R. Huang, 2019, Full-waveform inversion for salt: A coming of age: The Leading Edge, **38**, 204–213, doi: https://doi.org/10.1190/tle38030204.1.

Wang, P., Z. Zhang, Z. Wei, and R. Huang, 2018, A demigration-based reflection full-waveform inversion workflow: 88th Annual International Meeting, SEG, Expanded Abstracts, 1138–1142, doi: https://doi.org/10.1190/segam2018-2997404.1.

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, SEG, Expanded Abstracts, 1089–1093, doi: https://doi.org/10.1190/segam2014-0371.1.

Yang, Y., B. Engquist, J. Sun, and B. Hamfeldt, 2018, Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion: Geophysics, **83**, no. 1, R43–R62, doi: https://doi.org/10.1190/geo2016-0663.1.

Zhang, Z., J. Mei, F. Lin, R. Huang, and P. Wang, 2018, Correcting for salt misinterpretation with full-waveform inversion: 88th Annual International Meeting, SEG, Expanded Abstracts, 1143–1147, doi: https://doi.org/10.1190/segam2018-2997711.1.