Full waveform inversion with an exponentially-encoded optimal transport norm

*Lingyun Qiu *, Jaime Ramos-Martínez and Alejandro Valenciano, PGS; Yunan Yang and Björn Engquist, University of Texas at Austin*

Summary

Full waveform inversion (FWI) with L^2 norm objective function often suffers from cycle skipping that causes the solution to be trapped in a local minimum, usually far from the true model. We introduce a new norm based on the optimal transport theory for measuring the data mismatch to overcome this problem. The new solution uses an exponential encoding scheme and enhances the phase information when compared with the conventional L^2norm. The adjoint source is calculated trace-wise based on the 1D Wasserstein distance. It uses an explicit solution of the optimal transport over the real line. It results in an efficient implementation with a computational complexity of the adjoint source proportional to the number of shots, receivers and the length of recording time. We demonstrate the effectiveness of our solution by using the Marmousi model. A second example, using the BP 2004 velocity benchmark model, illustrates the benefit of the combination of the new norm and Total Variation (TV) regularization.

Introduction

FWI is formulated as a nonlinear inverse problem matching modeled data to the recorded field data (Tarantola, 1984). Usually, a least-square objective function is used for measuring the data misfit. This misfit is minimized with respect to model parameter and the model update is computed using the adjoint state method. FWI can produce high-resolution models of the subsurface when compared to ray-based methods. Due to the large scale of the problem, local rather than global optimization methods are mandatory. However, FWI is often an ill posed problem due to the band-limited nature of the seismic data and the limitations of the acquisition geometries. Furthermore, the non-convexity resulting from the least-square objective function causes the local minima, i.e., cycle-skipping problem, especially with data lacking low frequency information.

It is well known that the least-square formulation of FWI tends to produce many local minima. This is because only the pointwise amplitude difference is measured with L^2 norm while the phase or travel-time information embedded in the data is more critical for the inversion. There are different approaches proposed to capture the travel-time difference, such as dynamic time warping and convolution based methods. This information is used in order to convexify the objective function or enlarge the true solution valley. In this direction, we mention the works in (Luo and Sava, 2011), (Ma and Hale, 2013) and (Warner and Guasch, 2014).

Recently, the Wasserstein distance has been proposed to replace the L^2 distance for the objective function in FWI (Engquist and Froese, 2014). The Wasserstein distance is a well-defined metric from the theory of optimal transport in mathematics. It was first brought up by Gaspard Monge in 1781 (Monge, 1781) and more recently by Kantorovich (Kantorovich, 1942) seeking the optimal cost of rearranging one density into the other, where the transportation cost per unit mass is the Euclidean distance or Manhattan distance.

Wasserstein distance has the ability to consider both phase shifts and amplitude differences It has been demonstrated in (Engquist, Froese and Yang, 2016) that W^2 bears some advantageous mathematical properties, such as convexity with respect to shift and dilation and insensitivity to noise. In (Yang Engquist, Sun and Froese 2016), W^2 on 2D data is applied to FWI on synthetic benchmark models. The calculation of the corresponding adjoint source requires solving a Monge–Ampère equation that can be computationally demanding. Another popular optimal transport metric used for FWI is the 1-Wasserstein distance (W^1), approximated by the Kantorovich Rubinstein (KR) norm (Métivier, et al, 2016). For this metric the transport map is not unique. The KR norm doesn't require data to be positive and mass preserved. Therefore it can be directly applied to the seismic data without transferring them into probability density function (pdf). Both analysis and numerical results shows the potential of FWI with optimal transport to mitigate cycle-skipping problem.

The Wasserstein metric is designed to measure the distance between two pdfs. Thus, non-negativeness and unit mass are desired for the input. But, oscillation and sign-change are typical features of the seismic data. Therefore, we need a misfit function that takes the global features of data into consideration and is robust to periodicity and sign-change. Since seismic data are not naturally positive, a proper normalization method is the key to Wasserstein distance based inversion. Some previous methods may lead to non-differentiable misfit function and are not compatible with adjoint-state method, or lose information of original data during the normalization.

Here, we address the issue of how to transform seismic data into pdfs. The new solution uses an exponential encoding scheme and enhances the phase information when compared with the conventional L^2 norm. The algorithm

uses of the 1D Wasserstein metric. As a result, the implementation of the adjoint source has the same order of computational complexity as of the conventional L^2 norm. We illustrate our method by using the Marmousi and the BP 2004 velocity benchmark models.

Exponentially-encoded Wassesrtein distance for seismic data

In this section, we define a procedure to transfer the seismic data into pdf-like data before we calculate the Wasserstein distance between them. Meanwhile, we also pursue to extract the phase information from the seismic data for computing Wasserstein distance. Seismic data are not naturally positive, which is a challenge to apply W^2 directly. Some previous methods such as comparing the positive and negative parts separately (Engquist and Froese, 2014) seem not be compatible with adjoint-state method. The linear transformation (Yang Engquist, Sun and Froese, 2016) may lose the global convexity that W^2 has for positive signals. Therefore a proper data normalization method is the key for inversion.

Suppose we have seismic data d, which has both positive and negative values. We let

$$\tilde{d} = e^{\alpha d}$$

where α is a prescribed positive constant to control the upper bound of the power for the numerical accuracy. Since the exponential function has the feature that it has much milder derivative on the negative half real axis, the above procedure treat the negative and positive part of the seismic data differently. At the same time, the processed data is non-negative. We apply this procedure to both the recorded data and simulation with the same constant. With an additional scaling, we turn the recorded data d and simulated data u into pdf-like functions $\tilde{\boldsymbol{d}}$ and $\tilde{\boldsymbol{u}}$. Therefore, we can apply the Wasserstein distance to measure their difference.

Intuitively, the above algorithm is nothing but an uneven encoding process. All the information in the positive part of the data is amplified and stored in $(\mathbf{1}, +\infty)$ and the information from the negative part is compressed in $(\mathbf{0}, \mathbf{1})$. In this way, the phase information is extracted mainly from the positive side of the seismic data for the FWI. This encoding process is invertible and Fréchet differentiable. Therefore, according to the chain rule, the only additional work is to multiply the adjoint source by

$$\frac{\partial \tilde{d}}{\partial d} = \alpha e^{\alpha d}.$$

FWI with this encoding process will be biased to match travel-time provided by the positive signal. The negative side is also needed, especially for FWI with reflection data. To make use of the phase information from the negative part of the data, we balance this uneven encoding by also taking into account the data reformed by the map

$$\tilde{d} = e^{-\alpha d}.$$

In practice, we perform the inversion in an alternative fashion. That is, we switch the data encoding process between $\tilde{d} = e^{\alpha d}$ and $\tilde{d} = e^{-\alpha d}$ every few iterations.

The corresponding objective function is constructed as

$$J = \sum_{shot} \sum_{receiver} W_2^2(\tilde{u}, \tilde{d}).$$

Since we only change the objective function, the corresponding modification for the conventional FWI is to use a new adjoint source. It can be computed as

$$\frac{\partial J}{\partial u} = \sum_{shot} \sum_{receiver} \left(\frac{\partial}{\partial \tilde{u}} W_2^2(\tilde{u}, \tilde{d}) \frac{\partial \tilde{u}}{\partial u} \right).$$

Note that $\tilde{d}$ and $\tilde{u}$ are 1D functions. We can take advantage of the explicit expression of the Wasserstein distance for distributions over the real line. In this way, the computational complexity for obtaining the adjoint source is $O(N_r N_s N_t)$, where N_r, N_s and N_t stand for the number of receivers, shots and time steps, respectively. In practice, we find that the additional computational time is very small compared with the conventional method to calculate the adjoint source, which is a subtraction with the same order of complexity $O(N_r N_s N_t)$.

The quadratic Wasserstein distance between two 1D pdfs p_0 and p_1 is defined as

$$\begin{aligned} W_2^2(p_0, p_1) &= \int_0^1 (f_0^{-1}(s) - f_1^{-1}(s))^2 ds \\ &= \int_0^1 (f_0^{-1}(f_1(t)) - t)^2 p_1(t) dt \end{aligned}$$

Here, f_0 and f_1 are the associated cumulative distribution functions (cdf) and $\cdot^{-1}$ stands for the pseudo-inverse defined as

$$f^{-1}(t) = \inf\{s \mid f(s) > t\}.$$

The Fréchet derivative with respect to p_1 is given by

$$\frac{\partial W_2^2(p_0, p_1)}{\partial p_1} = (f_0^{-1}(f_1(t)) - t)^2 + \int_t^1 2 \frac{\partial f_0^{-1}(x)}{\partial x} |_{x = f_1(s)} (f_0^{-1}(f_1(s)) - s) p_1(s) \, ds.$$

The above equality can be simplified using the inverse function theorem and we have that

$$\frac{\partial W_2^2(p_0, p_1)}{\partial p_1} = (f_0^{-1}(f_1(t)) - t)^2 + 2 \int_{f_0^{-1}(f_1(t))}^1 (s - f_1^{-1}(f_0(s))) \, ds.$$

Note that both f_0 and f_1 are monotonic increasing functions. Hence, $f_0^{-1}(f_1(t))$ and $f_1^{-1}(f_0(t))$ are computed in $O(N_t)$ operations and both are monotonic functions. Therefore, we can obtain the adjoint source for a single trace with $O(N_t)$ operations. Once the adjoint source is obtained, the rest of the inversion is the same as the conventional FWI.

Numerical experiments

We first investigated the use of our method on the Marmousi model (Figure 1a). The model contains many reflectors, steep dips, and strong velocity variations in both the lateral and the vertical direction. The velocity model is 9.2 km × 3.2 km. The synthetic data was created with a minimum frequency of 5 Hz (zero power) and 7 Hz full power. The sources and receivers are both uniformly distributed every 20 m at 40 m depth. The maximum recording time is 8 s. We randomly select 31 sources per iteration. The initial model (Figure 1b) is created by smoothing the true model using a Gaussian filter with 2 km correlation length. With this initial model, inversion with $\boldsymbol{L^2}$ objective function fails to provide a good reconstruction (Figure 1c) but the $\boldsymbol{W^2}$ gives a result closer to the true model (Figure 1d).

Next, we perform numerical test on the BP 2004 benchmark velocity model (Figure 2a) (Billette and Brandsberg-Dahl, 2005). The model is 28.5 km × 7.5 km and contains a salt body in the middle of the domain of interest. The synthetic data was created with a minimum frequency of 1 Hz (zero power) and 3 Hz full power. For the acquisition geometry, the sources are uniformly distributed every 40 m and the receivers are deployed every 40 m with a maximum offset of 20 km. Both source and receiver are located at 40m depth. With this long-offset setting, the maximum recording time of the data is set to 12 s. For efficiency purpose, a random selected 36 shots are used per iteration.

A heavily smoothed model (1.1 km correlation length) from the true model with the water layer fixed is used as the starting velocity model for FWI (Figure 2b). From this initial model, the conventional FWI with L^2 distance fails to recover the salt boundary (Figure 2c). As shown in Figure 2d, inversion with proposed algorithm produces better reconstruction. The salt body shallower than 7 km depth is well restored. Slices of initial model, true model, L^2 reconstructed model and W^2 reconstructed model at x=12 km are shown in Figure 3.

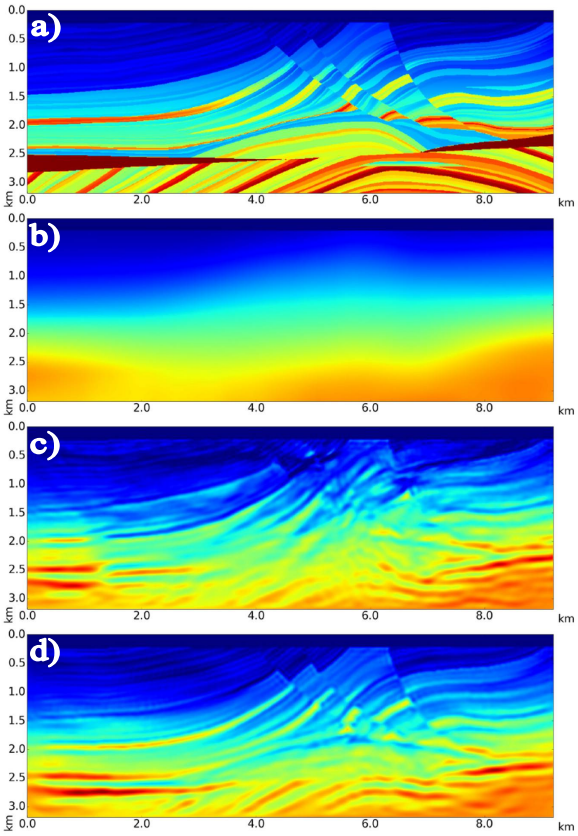

Figure 1: (a): True model, (b) Initial model, (c) FWI with $\boldsymbol{L^2}$ (d) FWI with $\boldsymbol{W^2}$.

In this work, we focus on measuring the difference in data space. Thus, no conditioning or stabilization procedure, such as smoothing on the gradient and regularization on the model, is applied to the inversion results shown in Figure 2 and 3.

The oscillatory noise in FWI can be efficiently removed using total variation type regularization (Qiu, et al., 2016). The regularization is necessary to stabilize the inversion and inject a priori information into the optimization. The extension of the proposed algorithm to incorporate TV regularization is straightforward. The inversion results are shown in Figure 4 and 5. The TV regularization helps to produce a blocky inverted model. But, from the slices view (Figure 5), it is clear that the FWI with $\boldsymbol{L^2}$ distance (blue curve) and TV regularization do not restore the salt boundary correctly. In contrast, the $\boldsymbol{W^2}$ model is close to the true model showing almost perfect sediment velocity and salt boundary reconstruction.

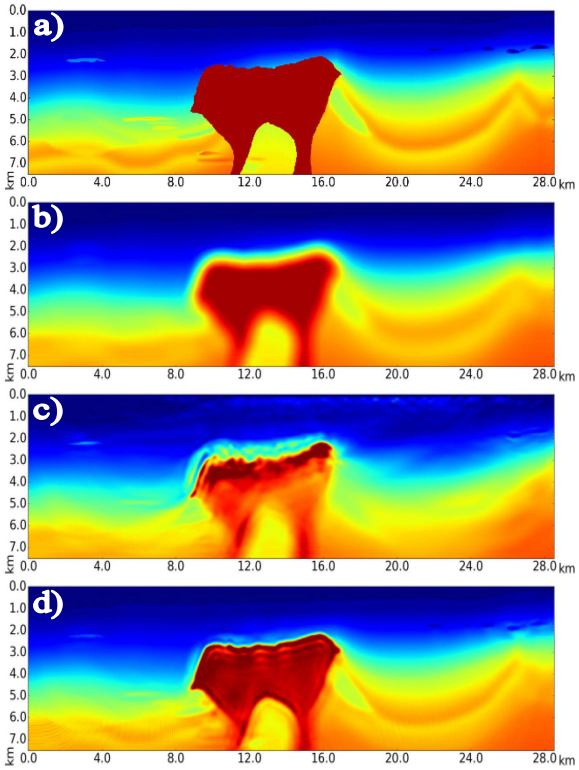

Figure 2: (a): True model, (b) Initial model, (c) FWI with $\boldsymbol{L^2}$ (d) FWI with $\boldsymbol{W^2}$.

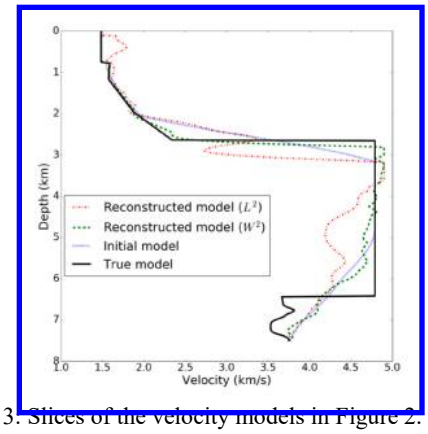

Figure 3: Slices of the velocity models in Figure 2.

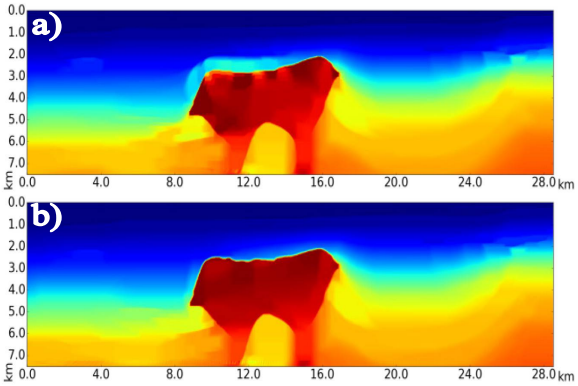

Figure 4: (a): $\boldsymbol{L^2}$ with TV regularization, (b): $\boldsymbol{W^2}$ with TV regularization

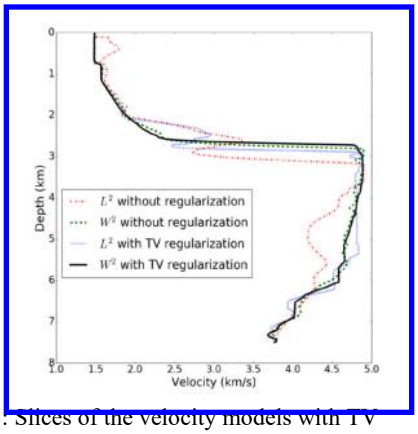

Figure 5: Slices of the velocity models with TV regularization in Figure 4.

Conclusions

The formulation of FWI with Wasserstein distance shows the potential to mitigate the cycle-skipping problem present in the $\boldsymbol{L^2}$ solution. We propose an exponential-encoding process to transfer the seismic data into pdf with emphasis on phase information. The adjoint source is calculated using the explicit solution of the optimal transport over the real line. All the efforts lead to an efficient and robust seismic inversion scheme. The numerical results demonstrate the advantages of the proposed algorithm. In the Marmousi example the new method allows the FWI to start from a heavily smoothed model with high frequency data and obtain a good result. The BP 2004 benchmark example shows how by combining the new norm with TV regularization the salt body velocity and boundaries can be reconstructed starting from a smooth model.

Aknowledgments

We thank PGS for permission to publish the results.

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2017 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

Billette, F. J., and S. Brandsberg-Dahl, 2005, The 2004 bp velocity benchmark: 67th Annual International Conference and Exhibition, EAGE, Extended Abstracts.

Ramos-Martinez, J., S. Crawley, S. Kelly, and B. Tsimelzon, 2011, Full-waveform inversion by pseudo-analytic extrapolation: 81st Annual International Meeting, SEG, Expanded Abstracts, http://dx.doi.org/10.1190/1.3627750.

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

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.

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

Yang, Y., B. Engquist, J. Sun, and B. D. Froese. Application of optimal transport and the quadratic wasserstein metric to fullwaveform inversion, arXiv preprint arXiv:1612.05075, 2016.

Metivier, L., R. Brossier, Q. Merigot, E. Oudet, and J. Virieux, 2016, Measuring the mis t between seismograms using an optimal transport distance: application to full waveform inversion: Geophysical Journal International, **205**, 345–377.

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

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.

Ma, Y., and D. Hale, 2013, Wave-equation refection traveltime inversion with dynamic warping and full-waveform inversion: Geophysics, 78, R223-R233, https://doi.org/10.1190/geo2013-0004.1.

Monge, G., 1781, Memoire sur la theorie des deblais et des remblais: Del'Imprimerie Royale.

Kantorovich, L. V., 1942, On the translocation of masses: In Dokl. Akad. Nauk SSSR, **37**, 199–201.

Qiu, L., N. Chemingui, Z. Zou, and A. Valenciano, 2016, Full-waveform inversion with steerable variation regularization: 86th Annual International Meeting, SEG, Expanded Abstracts, 1174–1178, http://dx.doi.org/10.1190/segam2016-13872436.1.