Deep-learning-based time shift estimation for Full Waveform Inversion

M. Alfarhan, F. Chen, G. Turkiyyah, D. Keyes, KAUST;*
I. Vasconcelos, Shearwater GeoServices;
M. Ravasi, presently Shearwater GeoServices, formerly KAUST.

Summary

Full Waveform Inversion (FWI) is a technique that leverages the discrepancy between modelled and observed seismic data to estimate a potentially high-resolution velocity model of the subsurface. However, due to the highly oscillatory nature of seismic waveforms, point-wise discrepancy measures are susceptible to cycle-skipping, particularly when starting from a poor initial velocity model. Over the years, various alternative misfit functions have been proposed, each with its own advantages and limitations. Dynamic Time Warping (DTW) is a widely used technique in signal processing for aligning two time series. Although a differentiable version of DTW has been recently developed, its application in gradient-based optimization faces challenges, including the presence of high-frequency artifacts in the adjoint source and the significant computational cost of gradient computation. In this work, we propose using a neural network to learn the time shift required to align a pair of time series in a supervised manner. The trained network is subsequently employed to compare traces from the observed and modelled data in FWI, offering a more computationally-efficient alternative to DTW. Moreover, as neural networks are inherently differentiable via back-propagation, the trained network can be seamlessly integrated into the misfit function of an FWI framework. We demonstrate the feasibility of this approach on the Chevron blind test dataset.

Deep-learning-based time shift estimation for Full Waveform Inversion

Introduction

Full Waveform Inversion (FWI) is a technique aimed at inverting seismic data for a high-resolution subsurface model by matching simulated and observed seismograms. However, due to the oscillatory nature of the seismic recordings, FWI is a highly nonlinear and ill-posed inverse problem that is prone to falling into local minima and thus failing to recover a geologically plausible subsurface model. This problem, also known as cycle-skipping, arises when the simulated and observed seismograms are separated by more than half a cycle.

Misfit functions focusing on the traveltime rather than amplitude mismatch between traces have been explored as a way to mitigate the cycle-skipping problem and include measures like cross-correlation (e.g., Van Leeuwen and Mulder, 2008, 2010), envelope (e.g., Bozdag et al., 2011), and optimal transport distance (e.g., Engquist et al., 2016), to name a few. Dynamic Time Warping (DTW – Sakoe and Chiba, 1978) has also been used in the context of FWI (Ma and Hale, 2013); however, its application involves high computational costs and hyperparameter tuning. Soft-DTW (Cuturi and Blondel, 2017), a differentiable variant of DTW, addresses some of these limitations, although it still requires careful formulation to avoid negative values and suboptimal minima (Chen et al., 2021).

Recent advancements in deep learning for time shift estimation have shown promise in seismic applications (Li and Abubakar, 2020). Building on this progress, we propose to use a neural network to estimate the time shift between two seismic traces and use it as an objective function within FWI; this approach offers differentiability, computational efficiency, and robustness against cycle-skipping, and provides results of comparable performance to DTW while being computationally more affordable.

Methodology

The objective of this work is to devise a neural network that predicts time shifts between two seismic traces, and embed it into the FWI workflow by minimizing the L2 norm of such time shifts. The process involves two stages: (1) a training stage, where the time shift mapping is learned in a semi-supervised manner, and (2) a deployment stage, where the trained network provides the time shift between modelledand observed traces, which is then used in the computation of the adjoint source for FWI. While we focus on a general traveltime-based objective function, in this study initially we apply the approach to refracted waves only – as is common practice for velocity model building. Thus, reflections are first masked from the observed seismic data to ensure compatibility with a smooth background velocity model, where reflections are weak or absent. At later inversion stages, reflection data is also included in the inversion.

Training dataset

The training dataset is created directly from the synthetic seismic traces modelled with the background velocity model to be used at the beginning of FWI. Such traces are shifted using a known time shift generated by summing multiple smooth Gaussian functions (i.e., $\tau(t) = \sum_{i=1:N} e^{-(t-\mu_i)/\sigma_i}$, where N, μ_i and σ_i are randomly sampled for user-defined distributions). Smoothness ensures stability when the predicted time shifts are used to compute the adjoint source for FWI. Moreover, to avoid misleading the network in areas where the seismic traces have zero amplitude, a windowing function is applied to the ground-truth time shift, ensuring it is nonzero only where an actual shift is present.

Network and loss functions

We adopt a 1D variant of the U-Net architecture (Ronneberger et al., 2015), which we modify to ensure smooth predictions. These adaptations include: (1) a smoothing layer after the last convolutional layer, (2) a kernel size of 11 × 11 (larger than what is commonly used in U-Net architectures), and (3) removal of the first skip connection to suppress high-frequency artifacts. The input of the network is composed by the concatenation of a reference $d(t)$ and its corresponding shifted trace $d^{\tau}(t) = d(t - \tau(t))$, and the output is the time shift $\tau(t)$. Training is performed in two phases. First, we train the network in a semi-supervised manner, meaning that we use a combination of mean absolute error (MAE) between true and predicted time shifts and mean squared error (MSE) between shifted (using true shift) and reconstructed (shifted using predicted shift) traces as our training objective function:

$$\mathcal{L} = \mathcal{L}_{MAE}\left(\tau(t), \tau_p(t)\right) + \mu\mathcal{L}_{MSE}\left(d\left(t - \tau_p(t)\right), d^{\tau}(t)\right), \tag{1}$$

where $\tau_p(t) = NN_\theta(d(t) \oplus d(t - \tau(t)))$ is the predicted time shift, and μ is a weighing factor to balance the losses. After performing our first phase of training, the network is fine-tuned using traces from the modelled data as reference traces and traces from observed data as shifted traces to account for other differences besides time shifts using the $\mathcal{L}_{MSE}$ loss.

Deployment: adjoint sources for FWI

Once trained, the network can predict time shifts that measure the misalignment between modelled and observed data. We define the FWI misfit as follows:

$$J(\mathbf{m}) = \frac{1}{2}\left\|\boldsymbol{\tau_p}(t)\right\|_2^2 = \sum_s \sum_r \frac{1}{2}\left\|NN_{\hat{\theta}}(d_{s,r}^{obs}(t) \oplus d_{s,r}(t))\right\|_2^2, \tag{2}$$

where $d_{s,r}^{obs}(t)$ and $d_{s,r}(t)$ are the observed and modeled seismic traces from a source s to a receiver r, respectively. The gradient of J w.r.t. the model parameters $\mathbf{m}$ can be written as:

$$\frac{\partial J}{\partial \mathbf{m}} = \lambda^T \frac{\partial J}{\partial \mathbf{d}}, \tag{3}$$

where $\partial J/\partial \mathbf{d}$ is obtained by back-propagating the L2 norm of the predicted time shift through the network (and represents the adjoint source), whilst we use λ^T to indicate the traditional FWI gradient computed via the adjoint state method (red arrow in Figure 1). Finally, the model is updated iteratively until convergence, aligning observed and modelled data in terms of their phase (traveltime) content.

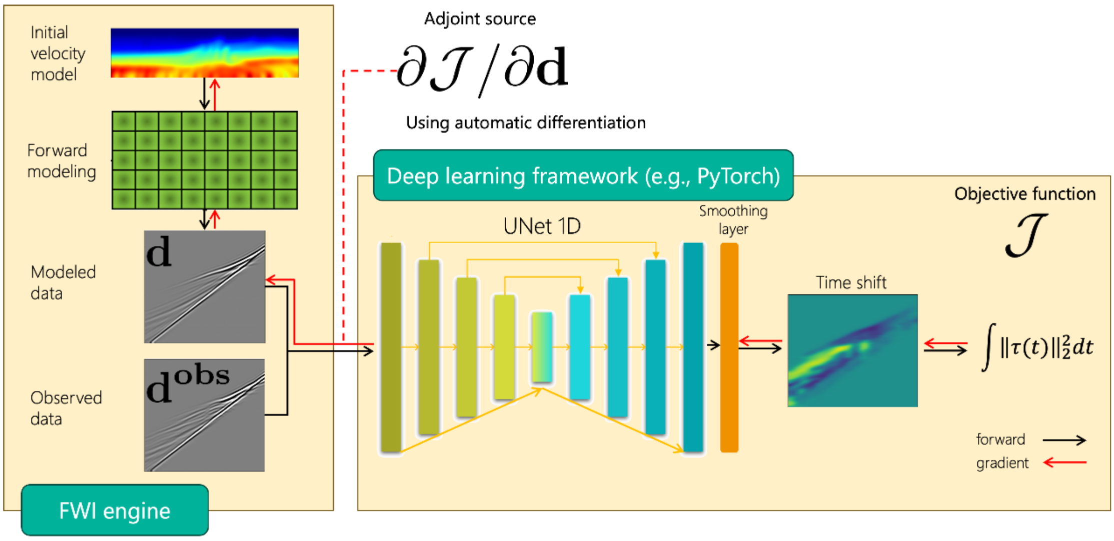

Figure 1 *Schematic representation of the proposed methodology where a neural-network-based time shift estimator is embedded in the FWI workflow.*

Numerical Examples

Chevron released a marine synthetic dataset as a benchmark for blind tests during the 2014 SEG FWI workshop. The dataset was modelled using the 2D isotropic wave equation with a free-surface boundary condition at the top. It features complex geology and is contaminated by significant noise at both the low and high-frequency ends of the spectrum. The data consists of 1,600 shots recorded by 321 hydrophones, each capturing 8 seconds of data with a 4 ms sampling interval. Chevron also provided a far-field wavelet, a single well log, and an initial velocity model (Figure 2(a)). The streamer acquisition geometry starts at a 1 km source-receiver offset, with both sources and receivers placed at a depth of 15 m. The receiver spacing is 25 m, and the maximum offset is 8 km. The model extends 6 km in depth and 47.75 km laterally. This dataset, along with the initial model, poses a significant challenge to conventional L2 norm-based FWI workflows due to the presence of cycle-skipping.

To test the robustness of our approach, we performed FWI on a subset of 256 shots, starting from the first shot and selecting every sixth shot up to the 1,531st shot. We compare our approach to the divergence form of soft-DTW (with a regularization parameter $\gamma = 100$, hereafter referred to simply as SoftDTW).

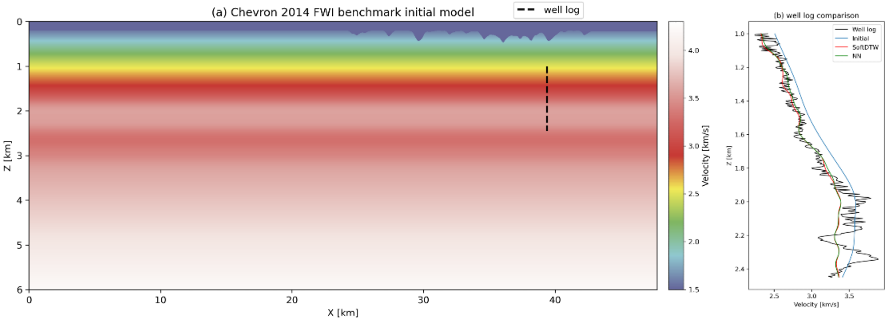

Figure 2 *The Chevron initial velocity model with the provided well log (trajectory at x = 39.375 km and z = [1.0, 2.45] km shown in black). (a) Initial velocity model, (b) velocity profile extracted along the well trajectory.*

In both cases, inversion is initially performed using only long-offset (e.g.,refraction) data, proceeding through three stages
with maximum frequencies of 4 Hz, 5 Hz, and 6 Hz, respectively. We then include reflections, progressing through three additional stages with maximum frequencies of 7 Hz, 10 Hz, and 15 Hz. Figure 3 compares the first FWI gradient computed at a maximum frequency of 4 Hz using both SoftDTW and our time-shift-NN-based misfit. The gradient from the proposed misfit closely resembles that from Soft-DTW in terms of shape and polarity, especially in the shallow part of the model. This similarity is crucial because the shallow region is most influenced by refractions, and it suggests that our predicted time shifts are guiding the modelled data toward the observed data correctly. Figure 4 shows the final velocity models after the complete inversion workflow, overlaid on the RTM images, for both misfit functions. Our FWI result appears very similar to that produced by SoftDTW. Furthermore, the well log comparison in Figure 2(b) shows that both methods have successfully recovered the correct velocities, with our approach being potentially more accurate in the shallower part of the model.

Our misfit function, however, offers two key advantages: it is inherently differentiable and notably more computationally efficient. First, the natural differentiability of neural networks makes integration with FWI straightforward in that adjoint sources are a by-product of the ML construct. Second, computing one gradient with our misfit function is about 60% faster than with SoftDTW (from the tslearn library) on an NVIDIA A100 GPU. Moreover, to manage SoftDTW's memory requirements, we had to downsample the data by a factor of five in time, which is not required in our approach. Assuming linear scalability, our approach provides a 92% reduction in computation time. Note that these numbers do not include training time for our approach, which is however negligible.

Conclusions

In this work, we proposed an NN-based-time-shift misfit function to compare modelled and observed datain FWI and demonstrated its effectiveness on the SEG 2014 Chevron synthetic dataset. Our results closely match those of SoftDTW in terms of accuracy, especially in the shallow part of the model, while offering significant computational and memory savings. More broadly, the proposed approach highlights the potential of neural-network-based misfit functions as efficient and robust alternatives to conventional methods for large-scale nonlinear inverse problems.

Acknowledgements

The authors thank KAUST and DeepWave sponsors for their support, and Shearwater Geoservices for direct contributions to the research. For computer time, this research
leveraged the resources of the Supercomputing Laboratory at KAUST in Thuwal, Saudi Arabia.

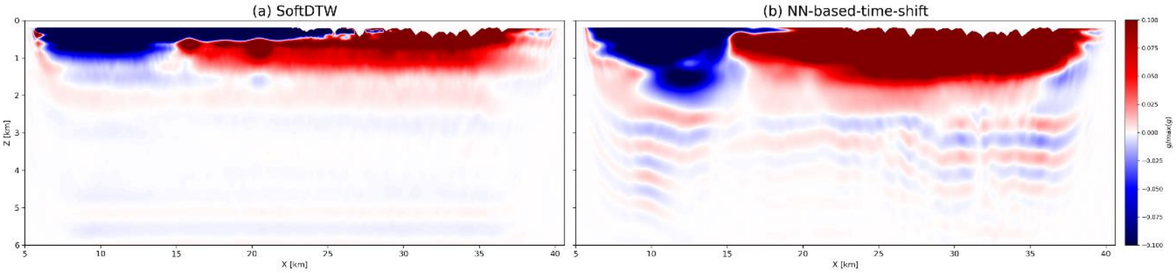

Figure 3 *The first iteration of FWI gradient, g, using (a) SoftDTW and (b) our approach.*

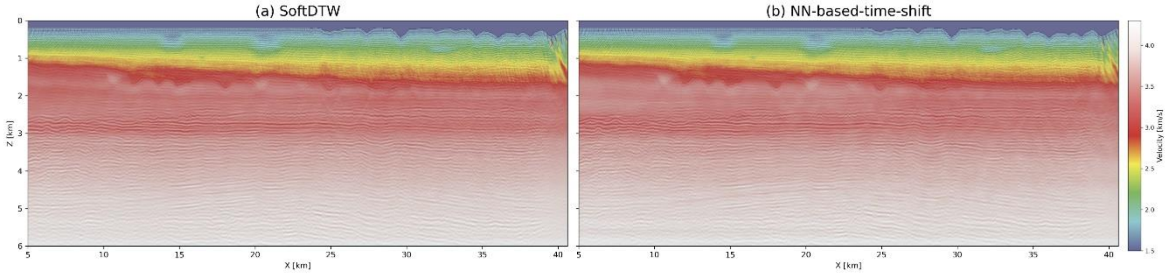

Figure 4 *Final updated velocity models following the complete inversion process, overlaid on RTM images, using (a) SoftDTW and (b) our approach.*

References

Bozdag, E., Trampert, J. and Tromp, J. [2011] Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. Geophysical Journal International, 185(2), 845–870.

Chen, F., Peter, D. and Ravasi, M. [2021] Misfit functions based on differentiable dynamic time warping for waveform inversion. In: SEG International Exposition and Annual Meeting. SEG,D011S029R002.

Cuturi, M. and Blondel, M. [2017] Soft-DTW: a differentiable loss function for time-series. In: International conference on machine learning. PMLR, 894–903.

Engquist, B., Froese, B.D. and Yang, Y. [2016] Optimal transport for seismic full waveform inversion. arXiv preprint arXiv:1602.01540.

Li, Z. and Abubakar, A. [2020] Complete sequence stratigraphy from seismic optical flow without human labeling. 1248–1252.

Ma, Y. and Hale, D. [2013] Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion. Geophysics, 78(6), R223–R233.

Ronneberger, O., Fischer, P. and Brox, T. [2015] U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, 234–241.

Sakoe, H. and Chiba, S. [1978] Dynamic programming algorithm optimization for spoken word recognition. IEEE transactions on acoustics, speech, and signal processing, 26(1), 43–49.

Van Leeuwen, T. and Mulder, W. [2008] Velocity analysis based on data correlation. Geophysical Prospecting, 56(6), 791–803.

Van Leeuwen, T. and Mulder, W.A. [2010] A correlation-based misfit criterion for wave-equation traveltime tomography. Geophysical Journal International, 182(3), 1383–1394