Neural network based low-frequency data extrapolation

Oleg Ovcharenko, Vladimir Kazei, Daniel Peter and Tariq Alkalifah*
King Abdullah University of Science and Technology

SUMMARY

Seismic full-waveform inversions is often hampered by a lack of low-frequency data, especially in the presence of salt bodies. Here, We propose a new way low frequency prediction from high frequencies using an artificial neural network (ANN). First, we train a neural network to predict low-frequency data from a sparse set of higher frequencies, using 10,000 random model realizations for a given source-receivers geometry. Second, we predict low-frequency spectra for the acoustic BP 2004 model. We compare predicted and true spectra for short- and far-offset data to demonstrate the feasibility of our neural network approach. The predicted low-frequency spectra could subsequently be used for early full-waveform inversion iterations in the multi-scale approach.

INTRODUCTION

Full-waveform inversion (FWI) is a technique that seeks to provide a seismic model that matches the observed data trace by trace by minimizing a data misfit (Warner et al., 2013). The conventional multi-scale approach (Bunks et al., 1995) incorporates successive inversions with gradually increasing frequencies. Inversions of low-frequency data create smooth and accurate starting models, whereas inversions of high-frequency data add more high-resolution details. While FWI efficiently constructs high-resolution estimates of the subsurface parameters without any limitations on the media complexity, its most significant drawback is that it converges to a local minimum, which is not necessarily coincident with the global one, when low frequencies are absent from the data.

There are numerous approaches that try to solve this problem by introducing advanced misfit functionals (Choi and Alkhalifah, 2015; Warner et al., 2016; Hu*, 2014), expanding the search space (van Leeuwen and Herrmann, 2013; Wu and Alkhalifah, 2016) and conditioning misfit gradients (Kazei et al., 2015, 2016; Alkhalifah, 2016). Recently Li and Demanet (2016) suggested extrapolating low frequencies for full-waveform inversions with robust initiation by using their event tracking technique (Li and Demanet, 2015). Each event is extrapolated to lower frequencies by using a linear phase relation. However, the event tracking involves some experimental tuning and requires different arrivals to be separable.

In this study, we propose to omit the step of event detection and instead construct an artifical neural network (ANN) that is trained to solve the problem of extrapolation. A feed-forward, neural network is a machine learning tool that can approximate an arbitrary non-linear function, having a sufficient number of neurons in a hidden layer topology (Ito, 1992). In the following, we investigate the feasibility of such a neural network approach for a realistic exploration scenario.

METHOD

Our aim is to predict low-frequency data based on band-limited, high-frequency recordings. The ANN approach taken in this investigation uses as input real- and imaginary-parts of the frequency-wavenumber (f-k) spectra from each receiver. The f-k domain discretization follows (Hu*, 2014) who showed that it is convenient to sample the spectrum with frequency steps equal to the targeted, low-frequency value. Additionally, the input data was normalized to have a zero mean with a standard deviation of 1 for the neural network. The target, output data are the true low-frequency amplitude values for each receiver.

The transformation of raw seismograms into the f-k domain reduces the dimensionality of the input, thus is better suited for our neural network approach. To make the neural network training more stable and faster, we filter out high-wavenumber data. The filtering is important as computational costs grow dramatically with increasing size of inputs and outputs, however with the disadvantage of losing some data information.

Recorded spectra encode the properties of underlying media. Thus, we try to find a non-linear operator that would represent the relation between the medium characteristics and the spectra at each receiver for a given source event location. This non-linear operator is the ANN described below.

Test case - BP model
The synthetic, true model chosen in this study is the BP 2004 salt model, resampled to be 6 x 33 km in size. It consists of 120 and 675 grid nodes in each direction, respectively. We intend to extrapolate 0.5 Hz data from 2-3.5 Hz data. The source spectrum is assumed to be known, yet the algorithm allows us also to create low frequencies with arbitrary source signatures in the low-frequency range.

Training set
We create 10,000 random models of the same size for the neural network training. Obviously, the true BP 2004 model is omitted from this training model set. These training models are generated by random perturbations of the model on a sparser grid. Figure 1 shows the BP model and a typical random model employed for training, together with the corresponding source-receivers configuration. Seismograms were recorded at the surface using a line of 201 geophones.

Neural network architecture
The neural network used in this study has two hidden layers with an optimal number of neurons in each layer selected according to (Huang, 2003). The input layer of our neural network (Figure 2) has the size of the reduced data for high frequencies (328 real values), and output layer has the size of data for a single low frequency (82 real values). Activation functions in both hidden layers were hyperbolic tangent functions (tanh), whereas the last output layer had linear functions.

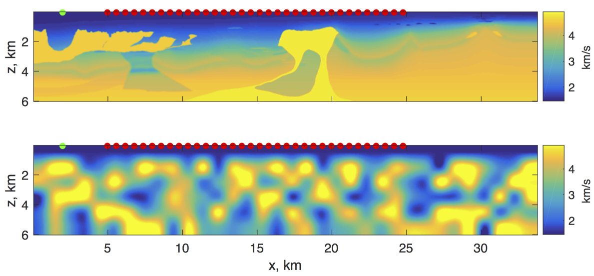

Figure 1: (top) BP 2004 salt model used in this study, representing the true model; (bottom) a random model utilized in the training set for the neural network. Source and receivers are indicated at the surface by green and red dots respectively.

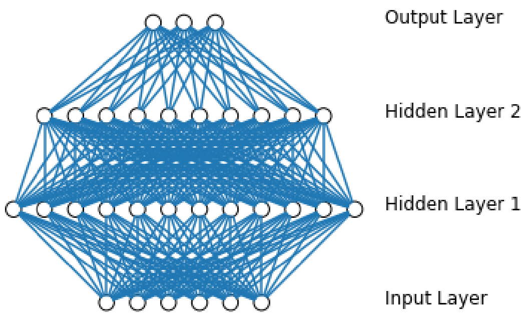

Figure 2: Schematic architecture of the neural network used. The optimal sizes of the first and second hidden layer are chosen to be 839 and 800 nodes

RESULTS

Training of a neural network is the process of adjusting the weights between layers. Errors between the true and predicted outputs are the measure of the training quality. Nesterov's optimization algorithm (Nesterov, 1983) was used to reduce this error on models used for training. The BP model was not included in the training set.

Duration of training on single CPU was about 8 hours, whereas the resulting extrapolation with the trained net could be done within a half of a second.

Low-frequency extrapolations

With the trained neural network, we predict the low-frequency spectra for all receivers using the true BP 2004 model and associated high-frequency data as input. Figure 3 shows the true and predicted target values in the frequency-wavenumber (f-k) domain. Notice that we left only 40 significant values. The extrapolated low-frequency spectrum (blue) follows reasonably well the true, modeled one (red).

The spectrum in frequency domain (f-x) of the data is shown in the Figure 4. True and predicted curves are close to each other. They do not coincide completely due to the dimensionality reduction we have made to accelerate the training. Another visible artifact are the wiggles at the far-offset (large receiver numbers). The latter effect has two causes: (i) the fundamental difficulty in predicting far-offset, low-frequency data from high-

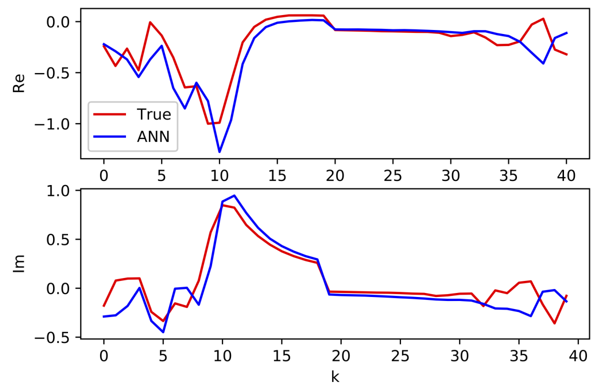

Figure 3: Predicted extrapolation with artificial neural network (blue) and true, modeled (red) low-frequency f-k spectra.

frequency short-offsets (Sirgue and Pratt, 2004; Kazei et al., 2013); (ii) Gibbs phenomenon due to the use of box filter in wavenumber domain to reduce the notorious data dimensionality. The later could be mitigated by using a smoother low-pass filter instead of a box filter.

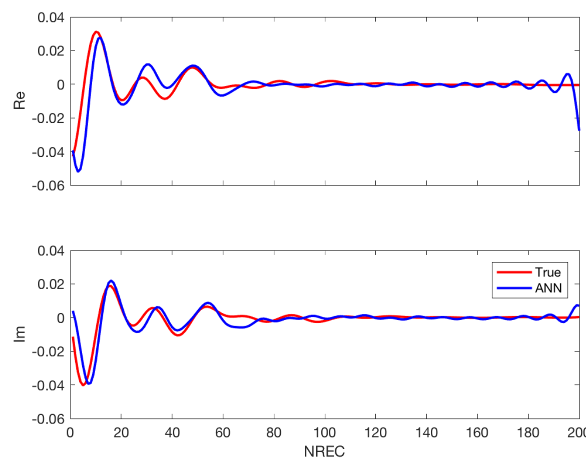

Figure 4: Predicted extrapolation with artificial neural network (blue) and true, modeled (red) low-frequency f-x spectra. Far-offset receivers (NREC $>$ 160) show a poor fit and could be muted for FWI purposes.

DISCUSSION AND CONCLUSIONS

We showed that the neural network trained with sparsely sampled f-k spectrum could, in principle, be used as a predictor of low-frequency data. Short-offset data seem to be better predicted with higher accuracy than far-offset data. In this study, we keep receivers fixed at certain locations for all training samples. This implies that the neural network has to be retrained for each new survey geometry. Future steps involve running full-waveform inversions with the predicted low-frequency data to define the limitations and benefits, estimating uncertainties and exploring further types of input/output data reductions.

REFERENCES

Alkhalifah, T., 2016, Full-model wavenumber inversion: An emphasis on the appropriate wavenumber continuation: GEOPHYSICS, **81**, R89–R98.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Choi, Y., and T. Alkhalifah, 2015, Unwrapped phase inversion with an exponential damping: Geophysics, **80**, R251–R264.

Hu*, W., 2014, Fwi without low frequency data-beat tone inversion, *in* SEG Technical Program Expanded Abstracts 2014: Society of Exploration Geophysicists, 1116–1120.

Huang, G.-B., 2003, Learning capability and storage capacity of two-hidden-layer feedforward networks: IEEE Transactions on Neural Networks, **14**, 274–281.

Ito, Y., 1992, Approximation of continuous functions on r d by linear combinations of shifted rotations of a sigmoid function with and without scaling: Neural Networks, **5**, 105–115.

Kazei, V., B. Kashtan, V. Troyan, and E. Tessmer, 2015, Pseudo-spectral full-waveform inversion: Seismic Technologies, 18–28.

Kazei, V., E. Tessmer, and T. Alkhalifah, 2016, Scattering angle-based filtering via extension in velocity, *in* SEG Technical Program Expanded Abstracts 2016: Society of Exploration Geophysicists, 1157–1162.

Kazei, V. V., B. M. Kashtan, V. N. Troyan, and W. A. Mulder, 2013, Spectral sensitivity analysis of FWI in a constant-gradient background velocity model: Presented at the 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013.

Li, Y. E., and L. Demanet, 2015, Phase and amplitude tracking for seismic event separation: Geophysics, **80**, WD59–WD72.

———, 2016, Full-waveform inversion with extrapolated low-frequency data: Geophysics, **81**, R339–R348.

Nesterov, Y., 1983, A method of solving a convex programming problem with convergence rate o (1/k2): Soviet Mathematics Doklady, 372–376.

Sirgue, L., and R. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667.

Warner, M., F. Herrmann, et al., 2016, Constrained waveform inversion-automatic salt flooding with inclusions: Presented at the 78th EAGE Conference and Exhibition 2016.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Štekl, L. Guasch, C. Win, et al., 2013, Anisotropic 3d full-waveform inversion: Geophysics, **78**, R59–R80.

Wu, Z., and T. Alkhalifah, 2016, The optimized gradient method for full waveform inversion and its spectral implementation: Geophysical Journal International, **205**, 1823–1831.