First-arrival waveform inversion using low-frequency regenerated data

Ken Xu and Cosmin Macesanu, SEIMAX Technologies

Summary

We propose a new approach to solve cycle skipping issues in full-waveform inversion (FWI) by regenerating low-frequency first arrivals. By using the regenerated first arrivals as the input data for the low-frequency FWI steps, we can provide an improved starting velocity model for the subsequent FWI iterations using the original band-limited data. Furthermore, the quality-control (QC) work of the first-break times can be done easily by comparing the regenerated first arrivals with the original ones. The method is robust since the regenerated first arrivals are just the synthetic data with full-frequency bandwidth and correct time shifts. Without complicated theory, only one conventional time-domain FWI algorithm is needed for both building the starting model, and final high-resolution velocity reconstruction.

Introduction

FWI has emerged as a very promising tool for building detailed seismic velocity models, which enhances the overall subsurface imaging quality. Conventional FWI updates the velocity model iteratively by minimizing the least-squares (LS) difference between the recorded and the predicted data. However, one fundamental challenge of conventional FWI is the local minimum issue caused by the cycle skipping between the input and predicted data.

To overcome this challenge, FWI requires sufficiently low frequencies. However, original seismic datasets from the field are always band-limited, thus very low frequencies (e.g., < 3 Hz) are usually unavailable. To overcome the cycle skipping in higher frequency bands, Shin and Cha (2008) proposed to extract low-wavenumber components from high-frequency seismic data, by distorting frequency information using Laplace-domain transformation. Hu (2014) also proposed a beat tone method. The distortions of frequency information may bring some potential risks in inversion, as discussed by Shin and Cha (2008). Most importantly, their schemes are based on the frequency-domain or Laplace-domain modeling, which may not be suitable for the large-scale computation requirements in production, mainly using parallelized time-domain modeling engines.

Without low frequencies, a highly accurate starting model is necessary for the success of FWI. This model is often built by first-arrival traveltime tomography (FATT) as discussed by Ravaut *et al.* (2004), and Xu and Greenhalgh (2010). Because the ray tracing method used in FATT is based on high-frequency approximations of the wavefield propagation, FATT has some inherent drawbacks as: (1) it may fail in low-velocity shadow zones, or highly varying velocity zones; (2) it has resolution limitation depending on shot/receiver intervals, grid size, and numbers of rays etc. To substitute ray tracing with full-wavefield modeling in tomography, some researchers, including Luo and Schuster (1991) and Choi and Alkhalifah, (2011) integrated the first-break traveltime information into the FWI adjoint-state framework. The time differences between the recorded and predicted data can be measured using a cross correlation function as discussed by Luo and Schuster (1991). However, the accuracy of the time difference estimation may vary from trace to trace due to changes in signal-to-noise ratio (SNR). These changes make it difficult to adjust the estimation coefficients or control the quality of the estimated times at each inversion iteration. In a different way, Choi and Tariq (2011) extracted the unwrapped phases to represent for the first break times in the frequency domain. Nevertheless, their algorithm needs explicit calculation of the imaginary part of the complex-valued frequency wavefield and its derivative with respect to each frequency. As a result, it is not suitable for time domain codes used for large-scale computations in production.

In this paper, we propose a new method to regenerate first-arrival waveforms with a full frequency band and with the same arrival times as the recorded ones. By using this synthetic data generated on a rough starting velocity model, the travel time information for the first arrivals in the original recorded data is transformed into the low-frequency phase information in the regenerated data. Thus, by using the low frequencies of the regenerated data for the initial FWI iterations, we can provide a better starting velocity model that will avoid cycle skipping for the subsequent FWI using the band-limited original data.

Theory

To regenerate a new dataset with low frequency information, we assume that the source wavelet is already known or estimated. We then establish a rough starting velocity model by using some simple preprocessing steps such as stack velocity analysis or well-log extrapolation. The synthetic dataset is generated using the same geometry as the original dataset; however the synthetic wavefields may have large amplitude and phase differences from the original data, if the initial model is not close to the true one. Phase differences for the first arrivals represent mainly the time shifts between the original and synthetic data. These time shifts can be estimated by a first-break time picking or by using a cross-correlation method. Once the time shifts are estimated, a new synthetic dataset can be regenerated by time-windowing and shifting the modeled first arrivals to the same times as the original first arrivals. Then, by comparing the original and regenerated first arrivals, we can easily quality-control (QC) the estimated time shifts, and either adjust the estimation parameters and redo the estimation, or eliminate the 'bad' estimated traces before inversion.

Since the same modeling engine is used for the regenerated dataset and predicted dataset in the initial inversion, the regenerated dataset is considered as a 'clean' dataset, with a known source wavelet, a full-frequency band, and no noise. By using the frequency bands in the regenerated dataset that are low enough to avoid cycle skipping, FWI generates a better starting velocity model by matching the time/phase differences between the regenerated and predicted first arrivals. Using the improved starting model to avoid cycle skipping, further FWI is conducted on the full wavefields of the original band-limited recorded dataset to reconstruct the high-resolution subsurface velocity models.

The workflow to implement the proposed method into FWI is

1) Model data using a rough starting velocity model;
2) Estimate the time shifts between the original and synthetic first arrivals by using a first-break time picking or a cross correlation method;
3) Time-shift the first arrivals of the synthetic dataset to the same locations as those of the original dataset;
4) Improve the starting velocity model by using low-frequency bands of the time-shifted synthetic first arrivals;
5) Further improve the starting velocity model by using the original first-arrivals (optional);
6) Conduct the final high-resolution multi-stage FWI using the full wavefields of the original dataset and the improved starting velocity model.

Algorithm

The modeling engine of our FWI algorithm is built with a high-precision, 2D/3D, isotropic or anisotropic, variable-density, acoustic-wave, finite-difference (FD)/pseudo spectral (PS) modeling code. The algorithm is parallelized with MPI to distribute shot simulations over cluster nodes, and is optimized at each node using shared-memory OpenMP. The order of the FD operator is 4 for time stepping, and 14 for the spatial grid. The top boundary of the modeling grid is treated as a free surface for the numerical tests, while at the other boundaries we implement absorbing boundary conditions.

The inversion scheme belongs to the conventional LS-FWI using the data residuals between the input and predicted data, and is implemented in the time domain. To reduce the amplitude mismatch and focus on the phase information, the amplitudes of the input data are normalized to the same level as those of the predicted data [as discussed by Warner (2014)]. In the inversion, the P-velocity is updated using the conjugate gradient method. The density model can be fixed or updated along with the P-velocity, using various velocity-density relations. For the later tests, we updated the density along with the P-velocity using one of their empirical relations. Time windowing and low-pass filtering can be applied, and adjusted, during different inversion stages, to reduce the nonlinearity of the inversion.

Examples

In our first example, we started our testing using a synthetic dataset created to approximate original data from the field by adding strong noise into the low-frequency (0-5 Hz) band of the input shot gathers, thus making it unsuitable for low-frequency FWI, and by adding lower levels of noise into the rest of the frequency band. The amplitude spectra of the original synthetic data and noisy synthetic data are displayed in Figure 1.

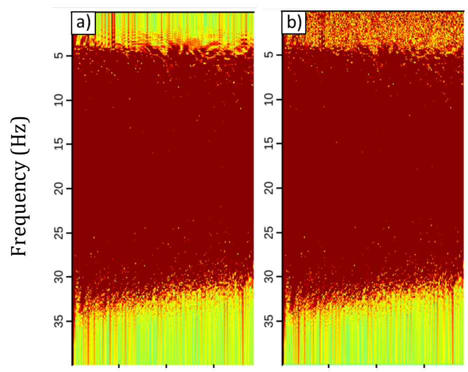

Figure 1: Amplitude spectra for (a) original synthetic data, and (b) noisy input data

The noisy synthetic data to be used as the input for later FWI tests is generated on a true acoustic Marmousi model, displayed in Figure 2(a). This model is also used in other robust time-domain inversion papers: Biondi and Almomin (2014); Warner (2014); Jiao *et al*., (2015). We used a 1D profile velocity model as a very rough starting model for FWI, displayed in Figure 2(b). We then ran conventional multi-stage FWI on the input data that lacked low frequencies (0-5 Hz), using the rough starting velocity model. Figure 2(c) demonstrates that FWI iterations using the higher frequency bands of the original input data (0-6 Hz up to 0-30 Hz) failed to converge, due to noise in the low-frequency band.

We then created a new synthetic dataset, using the rough starting velocity model in Figure 2(b). By time-windowing and applying a 0-6 Hz low-pass filter; we compared the first-arrival time/phase differences between the noisy input data, displayed in Figure 3(a) and the new synthetic data, displayed in Figure 3(b). Within the lowest available frequency band (e.g., 0-6 Hz) for the input data, their phase differences are still over half of cycle in the far offsets.

First-arrival waveform inversion

Using the estimated differences between the first arrivals, we time shifted the synthetic events in Figure 3(b) to the same time locations as those of the noisy input data in Figure 3(a) and created a regenerated synthetic dataset, with first arrivals displayed in Figure 3(c). The regenerated first arrivals carry the same, correct phase/time information as the input first arrivals displayed in Figure 3(a).

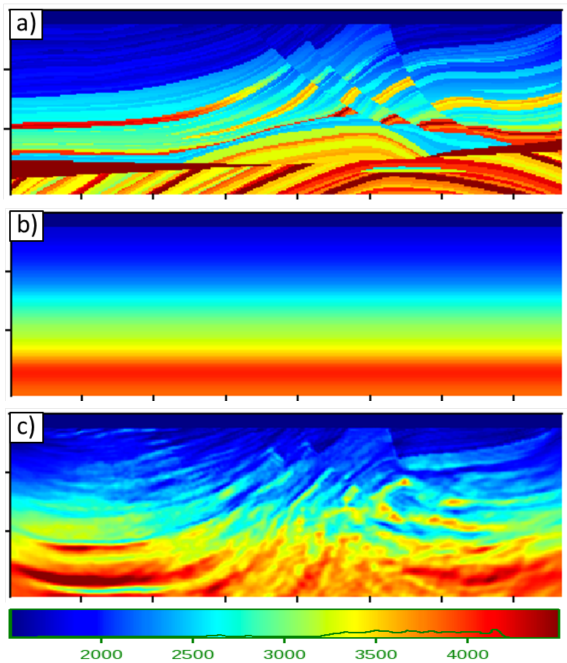

Figure 2: (a) True Marmousi model with noisy input data; (b) rough starting velocity model for FWI; (c) inversion result using a conventional LS FWI

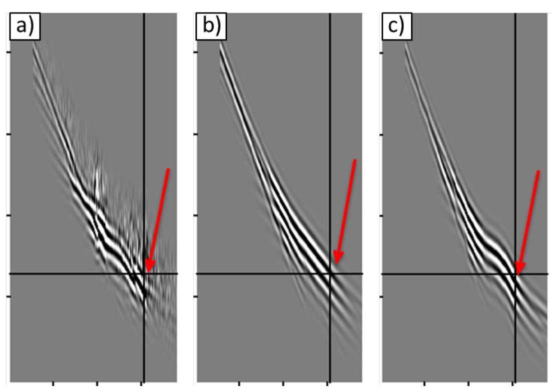

Figure 3: First arrivals at frequencies 0-6 Hz for the (a) noisy input data, (b) synthetic data generated with the velocity model in Figure 2b, and (c) regenerated synthetic data

After regenerating data with the full frequency band, any low frequency bands (e.g., 0-1.5 Hz) were then available for FWI, as displayed in Figure 4(b). Comparatively, the noisy input data in the same frequency band (0-1.5 Hz), is full of noise, as displayed in Figure 4(a).

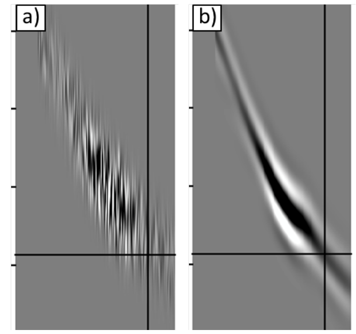

Figure 4: Low pass filtered data (0-1.5 Hz) for the (a) original, and (b) regenerated data

To test our theory, we began by using the regenerated synthetic dataset to conduct a multi-stage inversion within very low frequency bands; 10 iterations in 0-1.5 Hz bands, and a successive 10 iterations in 0-4 Hz bands. The velocity resulting from this procedure, displayed in Figure 5(a), is improved by matching the phases of the regenerated data.

We performed the inversion again, using the first improved starting model in Figure 5(a) and the first arrivals in 0-6 Hz frequency band from the original noisy data, depicted in Figure 3(a), to produce the second improved starting model displayed in Figure 5(b).

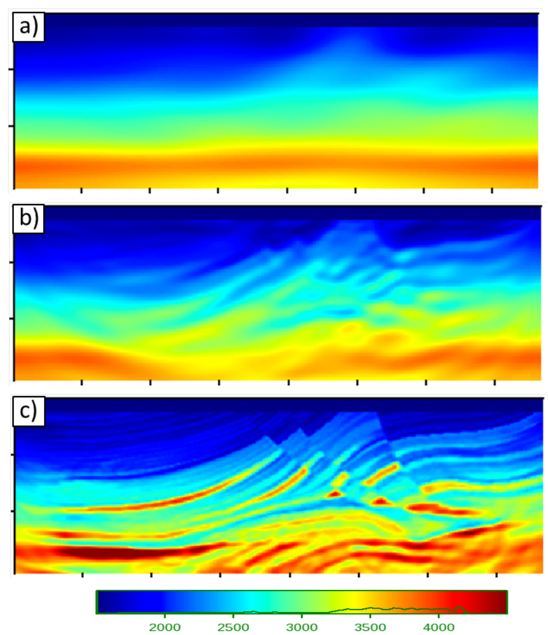

Figure 5: (a) the first improved starting model (0-4 Hz, regenerated data), (b) the second improved one (0-6 Hz, noisy input data), and (c) the final high-resolution result.

First-arrival waveform inversion

Since both the amplitude and phase information of the original first arrivals, depicted in Figure 3(a), are used in the inversion, Figure 5(b) has obvious higher resolution than Figure 5(a), which uses only the phase information of the regenerated data in the inversion.

Finally, we conducted a multi-stage FWI using the second starting model in Figure 5(b) and the higher frequency bands from 0-6 Hz to 0-30 Hz of the original full wavefields and successfully recovered the velocity model without cycle skipping, depicted in Figure 5(c).

In our second example, we used a highly realistic marine-environment seismic elastic model for FWI blind tests designed by Chevron in 2014. 1600 shot gathers, each with 321 traces, were generated by using an isotropic elastic modeling code with the free surface above, to simulate conventional 2D towed streamer lines with a maximum offset of 8000 meters.

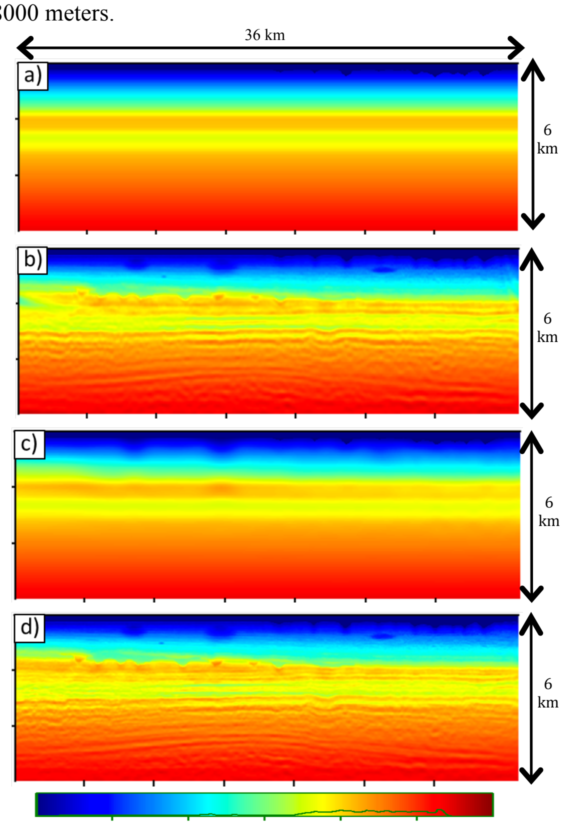

Figure 6: (a) Chevron rough starting velocity model; (b) conventional FWI with rough starting model; (c) improved starting model; (d) final FWI with improved starting model

To simulate the realistic environment, the data were frequency-band-limited without available low frequencies below 4~5 Hz, full of ambient noise, surface ghosts, and multiples. The relations between P-velocity, and density or S-velocity were also not deterministic.

The rough starting velocity model is displayed in Figure 6(a). We conducted a multi-stage conventional FWI from 0-5 Hz to 0-8 Hz using the Chevron rough starting model with results displayed on Figure 6(b). Since the starting model was far away from the true model, there were obvious artifacts related to cycle skipping [that can be seen on both sides of the model in Figure 6(b)], also observed by Warner (2015).

We created an improved starting model, displayed in Figure 6(c), using our low-frequency regeneration method and the low frequencies of the regenerated first arrivals below 5 Hz (e.g., 0-3 Hz), which was unavailable in the original datasets.

We conducted a final multi-stage FWI using the improved starting model from Figure 6(c) and the full wavefields of the original data (from 0-5Hz till 0-23 Hz), where we acquired a satisfactory high-resolution result, displayed in Figure 6(d), and is comparable to the results shown by Warner (2015).

Conclusions

In this paper, to avoid cycle skipping issues in FWI, the low-frequency regeneration method is proposed to convert the travel-time information of the first breaks into the low-frequency phase information in the regenerated first arrivals. Using the low frequencies of the regenerated data, FWI can reconstruct the low-wavenumber components of the subsurface velocity model, which is a better starting model for subsequent FWI updates using the original band-limited input data.

Since the regenerated first arrivals are actually synthetic data with the known wavelet, full frequency band, and free of noise, the method is robust and stable, similar to FATT, but without the ray-tracing theory's limitation. Most importantly, before the inversion, QC work can be easily done on the estimated time shifts, by comparing the original and regenerated first arrivals shot by shot. Finally, without complicated theory, only one conventional LS-FWI algorithm is needed for the whole process of starting-model building, and final high-resolution velocity reconstruction.

Acknowledgements

We thank our colleagues: Bill Starr, John Stevens, Joe Higginbotham, Faruq Akbar, Jun Ma, Kim Human, and Andrew Ryals for their fruitful discussions and suggestions on our research. We thank SEIMAX Technologies LP for permission for publication. We also thank Chevron for generating such a realistic and challenging dataset for FWI tests.

First-arrival waveform inversion using low-frequency regenerated data
Ken Xu and Cosmin Macesanu, SEIMAX Technologies

References

*Biondi, B., and Almomin, A., 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, no. 3, WA129-WA140.*

Choi, Y., and Alkhalifah, T., 2011, Frequency-domain waveform inversion using the unwrapped phase: 81st Annual International Meeting, SEG, Expanded Abstracts, 2576-2580.

Hu, W., 2014, FWI without low frequency data – beat tone inversion: 84th Annual International Meeting, SEG, Expanded Abstracts, 1116-1120.

*Luo, Y., and Schuster, G.T., 1991, Wave-equation traveltime inversion: Geophysics, **56**, no. 5, 645-653.*

Jiao, K., et al., 2015, Adjustive full waveform inversion: 85th Annual International Meeting, SEG, Expanded Abstracts, 1091-1095.

*Ravaut, C., et al., 2004, Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: Application to a thrust belt: Geophysical Journal International, **159**, 1032-1056.*

*Shin, C., and Cha, Y., 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, **173**, 922-931.*

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, SEG, Expanded Abstracts, 1089-1093.

Warner, M., and L. Guasch, 2015, Robust adaptive waveform inversion: 85th Annual International Meeting, SEG, Expanded Abstracts, 1059-1063.

*Xu, K., and Greenhalgh, S., 2010, Ore-body imaging by crosswell seismic waveform inversion: A case study from Kambalda, Western Australia: Journal of Applied Geophysics, **70**, 38-45.*