Adaptive data-selection elastic full waveform inversion

Zhen-dong Zhang* and Tariq Alkhalifah*

*King Abdullah University of Science and Technology.

SUMMARY

The multiscale inversion strategy is widely used to mitigate the cycle-skipping problem in full waveform inversion. There are many different approaches to implement the multiscale inversion and the widely used low-to-high frequency continuation is not applicable when the observed data lack low frequencies. As an alternative to multiple frequencies, offsets continuation is also able to suppress the cycle-skipping problem when the initial model is not perfect. We improve the multiscale strategy of offset-selection by introducing a local similarity criterion. Thus, we formulate an adaptive data-driven selection process that is better than conventional offset continuation approaches. The global-crosscorrelation objective function used here aims to maximize the similarity of two data sets instead of subtracting one from another and it is more consistent with the selection strategy. Besides, the crosscorrelation-based objective function is more sensitive to the phase information of the data and thus is more applicable to field data. We use a modified elastic Marmousi example to verify the effectiveness of the proposed method.

INTRODUCTION

Full waveform inversion (FWI) is instrumental to building a high-resolution model for the subsurface (Virieux and Operto, 2009). Conventional FWI intends to match the waveforms of the predicted data and those observed in the field. It, however, has inherent weaknesses in its application to field data. The L_2 norm objective function requires that the initial model should be kinematically reasonably accurate; otherwise, the inverse problem yields models corresponding to local minima when using a gradient-based optimization (Zhang et al., 2015, 2016). The multiscale inversion strategy implemented in different scenarios helps to avoid converging to a local minimum and it is usually implemented by gradually increasing the inversion frequency (Bunks et al., 1995). However, considering the lack of low frequencies and the unreliable seismic data amplitudes; the multi-frequency strategy is limited in its applications and the L_2 norm objective function is not practical.

Choi and Alkhalifah (2012) proposed a normalized global cross-correlation based FWI, which reduces the dependency on seismic amplitudes. It is more sensitive to the phase differences in the data, and thus, it is more immune to the ambient noise in the field data. However, the global crosscorrelation objective function still suffers from high nonlinearity, and the danger of converging to a local minimum when the initial model is far from the actual one. One intuitive remedy to the problem is to select parts of the data free from cycle-skipping in the inversion, which is referred to as multiscale inversion (Bunks et al., 1995; Bi and Lin, 2014; AlTheyab and Schuster, 2015).

A selection of frequencies from low to high is a widely used strategy in multiscale inversion. However, it is not applicable to the data lacking low frequencies. An alternative choice of multiscale inversion is selecting data corresponding to offsets, which are free of cycle-skipping. The crucial step of these approaches is the scheme used in selecting data for inversion in each iteration. Bi and Lin (2014) use the traveltime difference as a criterion to select the proper data. AlTheyab and Schuster (2015) gradually increases the offset of the data in the inversion. Although such approaches show promising inversion results, the L_2 norm objective function is inconsistent with such inversion strategies since increasing the selection range and the reduction in the data difference are competing with each other, and thus, can not be handled using current optimization schemes.

In our proposed approach, we use the local similarity to measure the differences between the predicted and the observed data and utilize the calculated similarity to adaptively select the data we will include in inversion. Thus, data that comply with the cycle skipping criteria for the updated model is automatically incorporated into inversion in each iteration. We use a global crosscorrelation to maximize the similarity of the predicted data and observed data. Both the selection range and the similarity should increase if the model is updated in the correct direction and thus the problem can be solved by existing optimization algorithms without modifications. This abstract is divided into four sections. After the introduction, the theory of the proposed inversion strategy is described. Then we use a synthetic elastic model to verify the effectiveness of the proposed method and draw the conclusions from the numerical test.

THEORY

Objective functions intend to measure the mismatch between the predicted and the observed data. One of the most intuitive measurements is the L_2 norm distance, which is given by

$$\phi(\mathbf{m}) = \|\mathbf{d}^p(\mathbf{m}) - \mathbf{d}^o(\mathbf{m})\|^2, \quad (1)$$

where ϕ measures the difference; $\mathbf{d}^p$ and $\mathbf{d}^o$ are predicted and observed data, respectively.

The inverse problem is constrained by the first-order elastic wave equation, which is given by (Virieux, 1986; Vigh et al., 2014; Zhang et al., 2017)

$$\begin{pmatrix} \rho \mathbf{I}_3 & 0 \\ 0 & \mathbf{C}^{-1} \end{pmatrix} \frac{\partial \Psi}{\partial t} - \begin{pmatrix} 0 & E^T \\ E & 0 \end{pmatrix} \Psi - \mathbf{f} = 0, \quad (2)$$

where $\Psi = (v_1, v_2, v_3, \sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6)$ is a vector containing three particle velocities and six stresses, $\mathbf{C}$ is the stiff-

ness matrix, E denotes space differentiation, and $\mathbf{f}$ is the source function.

Due to the oscillatory nature of seismic waves, the L_2 norm objective function suffers from cycle skipping when the mismatch between the predicted and observed data exceed a half-cycle. An intuitive remedy to this problem is to select parts of the data free of cycle-skipping and the process can be done adaptively. The proposed objective function is written as

$$J(\mathbf{m}) = - \sum_s \sum_r \mathbf{W} \hat{\mathbf{u}} \cdot \hat{\mathbf{d}}, \quad (3)$$

where $\hat{\mathbf{u}} = \frac{\mathbf{u}}{\|\mathbf{u}\|}$ and $\hat{\mathbf{d}} = \frac{\mathbf{d}}{\|\mathbf{d}\|}$ are normalized predicted and observed data, respectively. The indexes s and r correspond to the source and receiver locations, respectively, and $\mathbf{W}$ is the selection matrix, which will be explained later. The objective here is to maximize the similarity of the predicted and the observed data.

As discussed in the introduction, not all the data can be predicted with accurate kinematic information in practice. The conventional L_2 norm based FWI requires that the predicted and observed data have a maximum difference of half wavelength for each event. Otherwise, the adjoint source would be cycle-skipped. In our proposed method, we don't need to calculate the mismatch of each event because it is impractical to isolate each event. Instead, we use the local similarity measure to generate a weighting matrix, which emphasizes the most coherent events in the predicted and observed data. As we update the model, more of the data will fit the cycle-skipping criterion.

The local similarity proposed by (Fomel, 2007) was applied to compare the similarity of two images (i.e. the PP and PS image registration). Different from the global correlation, the local analysis produce a local correlation as a variable function that identifies local changes in the similarity of two signals. Each element of the similarity matrix is given by

$$w_s(t, r) = c_1 * c_2, \quad (4)$$

where t and r are indexes of time and receivers, respectively. $c_1 = \frac{\mathbf{S}^T \mathbf{d}}{\lambda^2 \mathbf{I} + \mathbf{S}(\mathbf{U}^T \mathbf{U} - \lambda^2 \mathbf{I})}$ and $c_2 = \frac{\mathbf{S} \mathbf{d}^T \mathbf{u}}{\lambda^2 \mathbf{I} + \mathbf{S}(\mathbf{D}^T \mathbf{D} - \lambda^2 \mathbf{I})}$. $\mathbf{U}$ and $\mathbf{D}$ are diagonal matrices composed from the elements of $\mathbf{d}$ and $\mathbf{u}$, respectively. The dividing operation is achieved by solving two least-square inverses.

The original definition of local similarity is not suitable for the problem because it ignores the polarity information of the two signals. For example, two signals with opposite polarity (fully cycle-skipped) have a similarity of 1 in this case. To overcome a potential fully cycle-skipping, we add a polarity detection in the selection matrix which is given by

$$w_p(t, r) = \begin{cases} 1, & \text{sign}(u) * \text{sign}(d) \geq 0 \\ 0, & \text{sign}(u) * \text{sign}(d) < 0 \end{cases},$$

and each element of the selection matrix, $\mathbf{W}$, is given by $w(t, r) = w_s(t, r) * w_p(t, r)$.

To obtain the gradient function of the proposed objective function, we take its derivative with respect to the model parameters as follows

ters as follows

$$\frac{\partial J}{\partial \mathbf{m}} = \sum_s \sum_r \frac{\partial \mathbf{u}}{\partial \mathbf{m}} \cdot \left(\frac{\mathbf{W}}{\|\mathbf{u}\|} (\hat{\mathbf{u}} (\hat{\mathbf{u}} \cdot \hat{\mathbf{d}}) - \hat{\mathbf{d}}) + \frac{\partial \mathbf{W}}{\partial \mathbf{u}} \right), \quad (5)$$

where $\frac{\partial \mathbf{W}}{\partial \mathbf{u}}$ can be approximated by $\frac{2.0 \mathbf{W}}{\mathbf{u}}$ and the divisions transforms products in practice.

The inverse problem here is nonlinear since the selection matrix is updated in each iteration. We cannot use L_2 norm based measurements in the inversion because the norm of $\mathbf{W}$ is increasing as we update the model and the data mismatch is reduced if the updating is in the right direction. In the proposed objective function (equation), the norms of $\mathbf{W}$ and $\hat{\mathbf{u}} \cdot \hat{\mathbf{d}}$ are both increasing as we update the model, and thus, allows for a consistent procedure treatable by current optimization methods.

The model is updated iteratively using the L-BFGS method (Liu and Nocedal, 1989), which is given by

$$\mathbf{m} = \mathbf{m}_0 - \lambda \mathbf{H}^{-1} \mathbf{g}, \quad (6)$$

where λ is the step length calculated by the standard line-search method. $\mathbf{H}$ is the approximated Hessian matrix.

NUMERICAL EXAMPLE

We use a modified Marmousi model to verify the effectiveness of the proposed method. The actual S-wave velocity is generated by setting $v_s = v_p / \sqrt{3} + 0.1 * (v_p - 2.4)$. The actual and initial velocity models are shown in Figures 1ab and 2ab, respectively. The source wavelet is a Ricker wavelet ($f_p = 5\text{Hz}$) with frequencies below 3 Hz muted, in which case a conventional L_2 norm based FWI cannot converge to the global minimum. Only one frequency band, 3-4 Hz, is used for the proposed inversion. The recorded data are two-component particle velocities. The staggered finite-difference method is implemented to solve the elastic wave equation. Initial models are 1D linearly increasing models, which are far from the exact one. We conduct a conventional global-correlation based FWI and the proposed FWI on the same data set for comparison. Figure 3 shows the estimated velocities using a conventional global correlation. There are limited updates due to the local minima constraint. Figure 4 shows the estimated velocities using the proposed method and it looks much better than the one from the conventional approach. We plot the vertical profiles in Figure 5 to highlight the accuracy. The estimated velocities using the proposed method better matches the actual ones than those estimated using the conventional global-correlation approach. There is a significant difference in the initial and actual velocity in many places as the black arrow indicates, which induces cycle-skipping even for later arrivals. The proposed method provides an automatic approach to capture related events through their similarity. As a quality control factor, we compare the data fitting histories of the conventional approach and the proposed approach as shown in Figure 6. The proposed approach has an overall slower but better convergence than the conventional approach. The data comparisons of the vertical component (Figure 7) and the horizontal component (Figure 8) also complements our claim of

the effectiveness of the proposed method. At last, we use the estimated velocities from the proposed method as initials for a conventional L_2 norm based elastic FWI and obtain the results in Figure 9. The final results are close to the actual models and two low-velocity zones have been recovered as the arrows indicate.

Figure 1: The actual models. a) v_p , b) v_s . There are two low velocity zones in v_p ; $v_s = v_p/\sqrt{3} + 0.1 * (v_p - 2.4)$.

Figure 2: The initial models. a) v_p , b) v_s . They're linearly increasing models and $v_s = v_p/\sqrt{3}$.

CONCLUSIONS

We proposed a data-selection elastic full waveform inversion based on local similarity criterion. A weighting function applied to the data is developed at each iteration by measuring the local similarity between the predicted and observed data. We also keep track of the sign difference to avoid high similarity of opposite polarity signals, and assign to them zero weight. We use the global correlation objective function for the FWI as it is consistent with the weighting criteria. The inverse problem in this case is nonlinear since the weighting function is updated in each iteration. It has a relatively lower convergence rate at the early iterations, but achieves better convergence overall. The elastic numerical example shows that the proposed method can provide an acceptable initial model for conventional L_2 norm based full waveform inversion. Although the

Figure 3: The estimated models using a conventional global crosscorrelation. a) v_p , b) v_s .

Figure 4: The estimated models using the proposed approach. a) v_p , b) v_s . They're used as initial models for a conventional L_2 norm based FWI.

Figure 5: The vertical profiles in the middle of the models. a) v_p , b) v_s . The estimated P-wave velocity using the proposed method has a better match with the actual one. The arrows indicate where the conventional global crosscorrelation based objective function fail.

Figure 7: The data comparison, vertical component. a) Observed data, b) predicted data from the initial model and c) predicted data from the estimated model using the proposed method.

Figure 8: The data comparison, horizontal component. a) Observed data, b) predicted data from the initial model and c) predicted data from the estimated model using the proposed method.

Figure 6: The normalized misfit versus iteration. The proposed method has a better convergence than the conventional one. The inversion stops when there are no updates.

Figure 9: The conventional L_2 norm based FWI using the velocities from the proposed method. a) v_p , b) v_s . Two low velocity zones have been recovered in the estimated v_p as the arrows indicate.

proposed method can suppress the cycle-skipping problem, it is not fully immune to cycle skipping. In the extreme case of an initial model far from the true one, we might end up with limited coherent events needed to calculate an effective gradient.

ACKNOWLEDGMENTS

We thank KAUST for its support and specifically the seismic wave analysis group members for their valuable insights. For computer time, this research used the resources of the Supercomputing Laboratory at King Abdullah University of Science & Technology (KAUST) in Thuwal, Saudi Arabia.

REFERENCES

- AlTheyab, A., and G. T. Schuster, 2015, Inverting reflections using full-waveform inversion with inaccurate starting models: 85th Annual International Meeting, SEG, Expanded Abstracts, 1148–1153, <https://doi.org/10.1190/segam2015-5849854.1>.
- Bi, H., and T. Lin, 2014, Impact of adaptive data selection on full waveform inversion: 84th Annual International Meeting, SEG, Expanded Abstracts, 1094–1098, <https://doi.org/10.1190/segam2014-0310.1>.
- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**, 1457–1473, <https://doi.org/10.1190/1.1443880>.
- Choi, Y., and T. Alkhalifah, 2012, Application of multi-source waveform inversion to marine streamer data using the global correlation norm: *Geophysical Prospecting*, **60**, 748–758, <https://doi.org/10.1111/j.1365-2478.2012.01079.x>.
- Fomel, S., 2007, Local seismic attributes: *Geophysics*, **72**, no. 3, A29–A33, <https://doi.org/10.1190/1.2437573>.
- Liu, D. C., and J. Nocedal, 1989, On the limited memory bfgs method for large scale optimization: *Mathematical programming*, **45**, 503–528, <https://doi.org/10.1007/BF01589116>.
- Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application using multicomponent measurements of seismic data collection: *Geophysics*, **79**, no. 2, R63–R77, <https://doi.org/10.1190/geo2013-0055.1>.
- Virieux, J., 1986, P-sv wave propagation in heterogeneous media: Velocity-stress finite-difference method: *Geophysics*, **51**, 889–901, <https://doi.org/10.1190/1.1442147>.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26, <https://doi.org/10.1190/1.3238367>.
- Zhang, Z., Y. Liu, and G. T. Schuster, 2015, Wave equation inversion of skeletonized surfacewaves: 85th Annual International Meeting, SEG, Expanded Abstracts, 2391–2395, <https://doi.org/10.1190/segam2015-5805253.1>.
- Zhang, Z.-d., T. Alkhalifah, and E. Z. Naeini, 2017, Multiparameter elastic full waveform inversion with facies constraints: 87th Annual International Meeting, SEG, Expanded Abstracts, 1551–1555, <https://doi.org/10.1190/segam2017-17672943.1>.
- Zhang, Z.-d., G. Schuster, Y. Liu, S. M. Hanafy, and J. Li, 2016, Wave equation dispersion inversion using a difference approximation to the dispersion-curve misfit gradient: *Journal of Applied Geophysics*, **133**, 9–15, <https://doi.org/10.1016/j.jappgeo.2016.07.019>.