Selective data extension for full-waveform inversion: An efficient solution for cycle skipping

<table>
<thead>
<tr>
<th>Item Type</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>Wu, Zedong; Alkhalifah, Tariq Ali</td>
</tr>
<tr>
<td>Eprint version</td>
<td>Publisher’s Version/PDF</td>
</tr>
<tr>
<td>DOI</td>
<td>10.1190/GEO2016-0649.1</td>
</tr>
<tr>
<td>Publisher</td>
<td>Society of Exploration Geophysicists</td>
</tr>
<tr>
<td>Journal</td>
<td>GEOPHYSICS</td>
</tr>
<tr>
<td>Rights</td>
<td>Archived with thanks to GEOPHYSICS</td>
</tr>
<tr>
<td>Download date</td>
<td>2023-11-01 11:16:51</td>
</tr>
<tr>
<td>Link to Item</td>
<td><a href="http://hdl.handle.net/10754/627338">http://hdl.handle.net/10754/627338</a></td>
</tr>
</tbody>
</table>
Selective data extension for full-waveform inversion: An efficient solution for cycle skipping

Zedong Wu and Tariq Alkhalifah

ABSTRACT

Standard full-waveform inversion (FWI) attempts to minimize the difference between observed and modeled data. However, this difference is obviously sensitive to the amplitude of observed data, which leads to difficulties because we often do not process data in absolute units and because we usually do not consider density variations, elastic effects, or more complicated physical phenomena. Global correlation methods can remove the amplitude influence for each trace and thus can mitigate such difficulties in some sense. However, this approach still suffers from the well-known cycle-skipping problem, leading to a flat objective function when observed and modeled data are not correlated well enough. We optimize based on maximizing not only the zero-lag global correlation but also time or space lags of the modeled data to circumvent the half-cycle limit. We use a weighting function that is maximum value at zero lag and decays away from zero lag to balance the role of the lags. The resulting objective function is less sensitive to the choice of the maximum lag allowed and has a wider region of convergence compared with standard FWI. Furthermore, we develop a selective function, which passes to the gradient calculation only positive correlations, to mitigate cycle skipping. Finally, the resulting algorithm has better convergence behavior than conventional methods. Application to the Marmousi model indicates that this method converges starting with a linearly increasing velocity model, even with data free of frequencies less than 3.5 Hz. Application to the SEG2014 data set demonstrates the potential of our method.

INTRODUCTION

Recently, full-waveform inversion (FWI), as a potential velocity model building tool, has gained a lot of momentum. FWI is iteratively capable of admitting high-resolution velocity models, provided that we start with a kinematically accurate (with respect to the minimum frequency available) initial guess of the model. However, when the initial velocity cannot accurately explain the kinematics of the wavefield within a half-cycle of the observed data, FWI usually converges to a local minimum instead of a global one, reflective of the cycle skip between the real and predicted data. These cycle-skipped predicted data result in an inaccurate velocity model, and usually this velocity model includes artifacts needed to produce reflections that fit the observed data.

To mitigate the cycle-skipping problem, many solutions have been proposed in recent years. One family of solutions suggest extending the model space (Symes, 2008; Biondi and Almomnin, 2014; Huang and Symes, 2015), with additional degrees of freedom that allows us to fit the data beyond the physical model limitations. A penalty on the unphysical nature of the model extension provides a path to correct the kinematics of the wavefield to make it suitable for FWI to converge. However, an extension of a model of any type usually forces the inversion to work with an extended model, which is relatively expensive. In addition, the penalty is applied in many forms, while allowing the basic measure of difference between the observed and modeled data to impose a classic FWI in the inversion. Thus, Warner and Guasch (2014) propose an adaptive FWI with the help of a Wiener filter. Another family of solutions to the cycle-skipping problem involves minimizing the lag of the maximum of the correlation between the observed and modeled data (Luo and Schuster, 1991; Chi et al., 2015). More generally, some researchers propose different methods to measure the phase difference between the modeled and observed data (Ma and Hale, 2013; Jiao et al., 2015; Yang et al., 2015). Métivier et al. (2016) propose to measure the difference between the observed and modeled data by an optimal transport distance. On the other hand, Bi and Lin (2014)
develop an adaptive data-selective method, which constrains the inversion to data that are not cycle skipped by a measure of the traveltime lag through crosscorrelation. Another group of methods is based on measuring the quality or the difference of extended image (Shen et al., 2003; Biondi and Symes, 2004; Sava and Biondi, 2004; Zhang and Schuster, 2013; Alkhalifah and Wu, 2017). For data without low enough frequency information, some researchers have proposed to generate artificial low frequencies by approximating the data in a transformed domain (Chin and Sha, 2008; Hu, 2014; Choi and Alkhalifah, 2015; Li and Demanet, 2016). For reflection-dominated data, Xu et al. (2012) and Zhou et al. (2012) develop a method based mainly on the work of Plessix et al. (1995) to invert for smooth velocity models using modeled reflected energy from an image. They refer to the method as reflected-waveform inversion (RWI). The idea is based on migration followed by demigration to predict the reflections. Because the demigration is obtained from the image, for an imperfect velocity, the modeled data from the image will have residuals, hopefully at far offsets. As a result, RWI inverts mainly the propagator (smooth) part of the model, such as migration-velocity analysis (MVA), without the need for extended images or angle gathers. Wang et al. (2013) implement the same inverts mainly the propagator (smooth) part of the model, such as image. They refer to the method as reflected-waveform inversion (RWI). The only difference is in the adjoint source, which makes the implementation, which they thought was necessary to avoid the nonlinearity caused by an incorrect image. Zhou et al. (2015) propose similar ideas that invert for velocity and impedance. In previous work, we proposed to invert for the background and perturbation simultaneously (Wu and Alkhalifah, 2015; Alkhalifah and Wu, 2016b), which can use diving waves, first-order reflection, and even multiscattering energy (Alkhalifah and Wu, 2016a) together. RWI can update along the waveshape of diving and reflected waves that can reduce the cycle-skipping problem in some way. However, RWI might still suffer from cycle-skipping problems inherited from the limitations in the objective function.

In this paper, we combine the data extension, which is less expensive than the model extension and the data-selective approach, to propose a new, efficient, and inexpensive solution for the cycle-skipping problem. We first extend the conventional zero-lag correlation between the observed and the modeled data in time (and possibly space) to keep the relative lag. We combine this extension with proper selective and weighting functions. Numerical examples show that the proposed method can have a much larger region of convergence (basin of attraction) than standard FWI. Because the extension is implemented in the data domain, the resulting algorithm has a similar gradient calculation and cost to that of standard FWI. The only difference is in the adjoint source, which makes the proposed method easy to implement.

THE OBJECTIVE FUNCTION BASED ON A SELECTIVE EXTENSION OF THE DATA

The standard waveform inversion can be formulated using the following optimization problem (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009):

\[
\min_{\mathbf{p}} J_0 = \sum_x \int_t \frac{1}{2} [p(x, t) - g(x, t)]^2 dt, \quad (1)
\]

s.t. \( \frac{1}{v^2} p_{tt} - \Delta p = f \), \quad (2)

in which \( v \) is the P-wave velocity, \( f \) is the source wavelet, \( g \) is the observed data, and \( p \) is the modeled wavefield corresponding to velocity \( v \). Here, we consider only a single-source case for simplicity. We only need to sum the contribution of the different source in the case of multiple sources. And the summation of \( x \) is over all the receiver locations. This classic objective function \( J_0 \) is highly sensitive to the amplitude of the wavefield, which is hard to accurately simulate considering our typical acoustic representation of the model and our ignorance of attenuation, among other shortcomings in the modeling process. An amplitude-independent objective function (Choi and Alkhalifah, 2012) is given by maximizing the crosscorrelation between the modeled and observed wavefields, which can be expressed as

\[
\min_{\mathbf{p}} J_1 = -\sum_x \frac{\int_t p(x, t) g(x, t) dt}{\sqrt{\int_t p(x, t)^2 dt \int_t g(x, t)^2 dt}}, \quad (3)
\]

The objective functions \( J_0 \) and \( J_1 \) are highly nonlinear with respect to the velocity model such that conventional local gradient-based methods often converge to local minima if the initial velocity is far from the exact one, there are not low enough frequencies in the data, or the maximum offset is not large enough. This is often attributed to the cycle skipping between predicted and observed data. An example of that is shown in Figure 1a, in which the red curve represents the observed data and the blue curve is the modeled data. The correlation between the red and blue curves equals zero because there is no intersection between the energy corresponding to the two events. In this case, gradient methods cannot improve the velocity because the objective function \( J_1 \) (Figure 1b) is flat around that velocity. To solve this problem and allow for an interaction between the modeled and observed data, it might be possible to move the observed data to the modeled data and correlate them, as shown in Figure 1c, which is the main objective of this paper.

A direct extension of objective function \( J_1 \) leads to

\[
\min_{\mathbf{p}} J_2 = -\sum_x \sum_{\tau, \mathbf{h}} W(\tau, \mathbf{h}) M(\zeta(p, g, \mathbf{x}, \tau, \mathbf{h})), \quad (4)
\]

in which \( W(\tau, \mathbf{h}) \) is a weighting function, \( M(\zeta) \) is a selective function, which we will discuss later, and the shifted crosscorrelation is defined as

\[
C(p, g, \mathbf{x}, \tau, \mathbf{h}) = \frac{\int_t p(x, t) g(x + \mathbf{h}, t + \tau) dt}{\sqrt{\int_t p(x, t)^2 dt \int_t g(x + \mathbf{h}, t + \tau)^2 dt}}, \quad (5)
\]

If the weighting and selective functions are chosen as

\[
W(\tau, \mathbf{h}) = -\left( \frac{|\tau|}{\max |\tau|} \right)^2 - \left( \frac{||\mathbf{h}||}{\max ||\mathbf{h}||} \right)^2, \quad \mathbf{M}(\zeta) = \zeta^2, \quad (6)
\]

the objective function \( J_2 \) becomes some kind of MVA in the data domain. However, this function reduces the comparison between observed and predicted data to a measure of lag in \( |\tau| \) or \( ||\mathbf{h}|| \), thus, providing traveltime-based smooth updates. Also, it is sensitive to the choice of maximum lag \( \mathbf{h} \) or \( \tau \) because the larger the lag, the bigger the weight in the objective function. Thus, alternatively, we suggest the following weighting function:
\[ W(r, h) = \left(1 - \frac{|r|}{\max |r|} \right)^2 \left(1 - \frac{|h|}{\max |h|} \right)^2, \] (7)

which enhances the wavefield comparison. In this case, the objective function maximizes the correlation, not only at zero lag, but also at nonzero lag. An example of this weighting function is shown in Figure 2. We can control the maximum shift by making the \( \max |r| \) and \( \max |h| \) depend on the quality of the initial velocity. In the case in which the modeled data are far from the observed data as shown in Figure 1a, we choose a large \( \max |r| \) (\( \max |h| \)). Of course, some other weight function, such as a Gaussian, might also be possible. In practical applications, we only need to shift over time \( \max |r| > 0 \) (\( \max |h| = 0 \)) or space \( \max |h| > 0 \) (\( \max |r| = 0 \)) to attain the required interaction between the observed and predicted data in spite of the cycle skip at zero lag. The two options of shifting schemes have their own advantages and disadvantages. The shift over space heavily relies on the geometry of the experiment. In the case of irregular geometry, the shift over space might be difficult to implement. However, in some cases, such as in frequency-domain inversions, the time-domain data are not available, and the shift over space may provide a viable alternative. Because we operate in the time domain, it is far more convenient and efficient to rely on the shift in time.

Nevertheless, we still cannot obtain pseudoglobal convergence because the correlation may admit negative values, which will result in negative contributions to the objective function, and thus, local minima. To enlarge the convergence region, we need to choose a special selective function to remove the negative contribution. Similar to Bi and Lin (2014), we choose the following selective function:

\[ M(\zeta) = \zeta, \quad \forall \zeta \geq 0, \quad M(\zeta) = 0, \quad \forall \zeta < 0. \] (8)

The selective function indicates that only the positive contribution to the objective function is admitted. As shown in Figure 3, it is a continuous function. However, the derivative of the objective is discontinuous.

A smoothed version of the selective function can be defined as

\[ M(\zeta) = \zeta^3, \quad \forall \zeta \geq 0, \quad M(\zeta) = 0, \quad \forall \zeta < 0. \] (9)

In this case, the second-order derivative of the selective function \( M(\zeta) \) is still continuous. In our numerical examples, we use the selective function 8 for simplicity.
THE GRADIENT CALCULATION AND SMOOTHING

To solve the optimization problem with a local-based optimization method, we derive the gradient of the objective function. We perturb the velocity \( v \) by \( \delta v \), so that

\[
\delta J_2 = -\sum_x \sum_{\tau,h} W(\tau, h) M'(C(p, g, x, \tau, h)) \delta C(p, g, x, \tau, h),
\]

where

\[
\delta C(p, g, x, \tau, h) = \frac{\int \delta p(x, t) g(x + h, t + \tau) dt}{\sqrt{\int p(x, t)^2 dt \int g(x + h, t + \tau)^2 dt}}
\]
the above terms that form the adjoint source is discontinuous, as shown in Figure 3. We need to apply some addition, the selective function $W(r, h)M'(C(p, g, x, r, h))$, given by

$$
\sum_{r,h} W(r, h)M'(C(p, g, x, r, h)),
$$

where $g(x + h, t + \tau)$

$$
\sqrt{\int p(x, t)^2 dt} \int g(x + h, t + \tau)^2 dt
$$

$$
- \frac{\int p(x, t)g(x + h, t + \tau)dt \int p(x, t)\delta p(x, t)dt}{\sqrt{\int p(x, t)^2 dt} \sqrt{\int g(x + h, t + \tau)^2 dt}}.
$$

As a result, the gradient is given by the adjoint-state method (Plessix, 2006) as follows:

$$
\nabla_r J_2 = - \frac{2}{v^3} \int p_n(x, t)\lambda(x, t)dt,
$$

where $\lambda$ is the back-propagated wavefield with $R(x, t)$ forming the adjoint source.

As we can see from the above gradient, the proposed method has the same cost as standard FWI for each iteration. We only need to execute one forward-modeling and one backward-modeling operation per iteration. The only difference is in the computation of the adjoint source. Due to the simple form of this objective function, it can be easily combined with other methods, such as RWI (Wu and Alkhalifah, 2015; Alkhalifah and Wu, 2016a, 2016b). In addition, the selective function $M(\zeta)$ is continuous and its derivative is discontinuous, as shown in Figure 3. We need to apply some smoothing to the obtained gradient, which is especially important for noisy data. Let us consider the popular Gaussian smoothing as our smoothing tool. Gaussian smoothing can be defined as a convolution operator as follows:

$$
Sf(x) = \int G(x - y)f(y)dy.
$$

in which the Gaussian kernel function is defined as

$$
G(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{|x|^2}{2\sigma^2}\right).
$$

With a bigger max $r$ or max $h$, the smoothing radius is bigger. This will result in a smoother gradient, which is consistent with the idea of multiscale inversion.

**NUMERICAL EXAMPLES**

To demonstrate that the new objective function can tackle the cycle-skipping issue, we start with a simple constant-velocity example. The exact velocity is $2$ km/s. The source is located at $0.1$ km in depth and in the middle of the model, laterally. The receiver is located at depth $3$ km. For our proposed method, we only shift the recorded data a max $|\tau| = 0.6$ s along the time axis. We do not shift in space for efficiency. The objective function ($J_2$) of the proposed method and the standard normalized correlation FWI ($J_1$) for a range of predicted constant velocities are shown in Figure 4a. We can clearly see that the proposed objective function has only one minimum at the global location, whereas standard FWI has more than three local minima. When the velocity is higher than $2.3$ km, the objective function of standard FWI is constant, independent of velocity. Also, we plot the observed (Figure 4b) and modeled data for a velocity equal to $4$ km/s (Figure 4c), and we overlay these two data in

![Figure 5](image-url)
Figure 4d. The two events do not overlap anywhere. In this case, even the adaptive data-selective method (Bi and Lin, 2014) will fail. This demonstrates the importance of the extension.

We now test the new objective function on the far more complicated Marmousi model. The source wavelet, which has no frequencies lower than 3.5 Hz, is shown in Figure 5a. The frequency spectrum of the source wavelet is shown in Figure 5b. The exact velocity is shown in Figure 6a. The observed data correspond to 39 shots, which are equally distributed on the surface. All grid points on the surface serve as receivers. We use the same wavelet for inversion, starting from a linearly increasing initial velocity model shown in Figure 6b. To invert the velocity with a multiscale approach, we choose the max τ sequentially as 0.45, 0.15, 0.075, and 0 s. For each stage of the inversion, the inverted model from the previous stage (previous τ) serves as the initial model. The inverted velocities with \( \max \tau = 0.45 \text{ s} \) (max \( \tau = 0.15 \text{ s} \), \( \max \tau = 0.075 \text{ s} \)) are shown in Figure 6c–6e. Finally, we apply the standard normalized correlation FWI (\( J_1 \)) by setting \( \max \tau = 0 \) followed by the objective function \( J_0 \), and the resulting inverted velocity is shown in Figure 6f. As we can see from these figures, the inverted velocity shows higher resolution as we choose a smaller \( \max \tau \). The proposed algorithm can obtain convergence, even starting from a linearly increasing initial velocity and no data less than 3.5 Hz. To show the accuracy of the inverted model, the velocity profiles at 3, 5, and 7 km are shown in Figure 7a–7c. Figure 8a shows the observed data. Figure 8b shows the modeled data generated by the initial velocity. Because of the linearly increasing initial velocity, the modeled data generated by initial velocity do not admit any reflections. Figure 8c and 8d shows the difference between the modeled data using the initial (inverted) velocity and observed data. As we can see, the inverted velocity is a much better approximation than initial velocity.

Our last example is the SEG2014 blind inversion challenge synthetic data set. It was distributed as a blind test to apply FWI. The data set includes 1600 shots records with 25 m shot sampling at the depth of 15 m. Each shot has 321 receivers with 25 m sampling at depth of 15 m. The data represent a marine acquisition, and they are generated with an isotropic elastic-wave equation. The maximum recording time is 8 s. The data have a low signal-to-noise ratio of less than 3 Hz and strong noise even in the 3–5 Hz range. To increase the signal-to-noise ratio and reduce the boundary effect near
the source, we convert the data into 641 receivers located on both sides of each shot, relying on the principle of reciprocity.

To reduce the computational cost associated with the inversion, we use an approach suggested by Díaz and Guitton (2011) and Reker et al. (2014) and divide the data set into eight groups. We then invert one group after another. An initial velocity was provided and shown in Figure 9a. We use the method proposed in Hicks (2002) to address the fact that shots and receivers are not on the regular grid points of the model. To avoid the large noise at low frequencies, we apply the proposed approach on the frequency bands of 0–6, 0–10, and 0–18 Hz, sequentially. For each frequency range, we first apply the proposed method with max $\tau = 0.45$, 0.15, 0.075 s and then we apply standard FWI. The frequency strategy is based on the approach suggested by Sirgue and Pratt (2004). Also, we apply Gaussian smoothing to the gradient with respect to $v$ with the smoothing radius defined according to equation 18 to reduce the artifacts caused by noise and multiple and elastic effects. The inverted $v$ is shown in Figure 9b. Figure 10 shows the well logs at 39,375 m from 1000 to 2500 m depth for comparison. We can see from Figure 10 that the initial velocity (pink line) is far away from the exact velocity (red line). Figure 11a–11c shows a data comparison between the modeled (Figure 11b and 11c) using the initial inverted velocity, and observed data (Figure 11a). We can see that the predicted data are similar to the observed data. To show the accuracy of the inverted model, we define the quality factor as

$$\text{Correlation}_s = \frac{\sum_r \sum_t s_r(x_r,t)g_s(x_r,t)}{\sqrt{\sum_r \sum_t s_r(x_r,t)^2 \sum_r \sum_t g_s(x_r,t)^2}}$$

(19)

where $g_s(x_r,t)$, $s_r(x_r,t)$ are the observed and modeled wavefields, respectively, from the shot $s$ and receiver $r$, and $x_s, x_r$ are their relative locations. If the inverted model can produce exactly the same data as observed one, $\text{Correlation}_s = 1$. We show the correlation for different shots between the data generated by our inverted model and the observed data. Thus, Figure 12 indicates that the inverted model can predict reasonably accurate approximation data to the observed data. Figure 13a and 13b shows the reverse time migration result with

![Figure 8](image_url)

Figure 8. (a) Observed data. (b) Modeled data with the initial velocity. (c) The difference between the observed and modeled data by the initial velocity. (d) The difference between the observed and modeled data by the inverted velocity.
the initial and inverted velocities. The image in Figure 13b is more continuous and better focused. To show the quality of the resulted reverse time migration image, we compute angle gathers using the space-shift imaging condition (Sava et al., 2005). The resulting angle

Figure 9. (a) Initial and (b) inverted velocity models.

Figure 10. Well-log comparison (pink curve, initial velocity; red curve, exact velocity; green curve, inverted \( v \) with our proposed method; and blue curve: inverted \( v \)).

Figure 11. Data comparison. (a) Observed data. (b) Modeled data by initial velocity. (c) Modeled data by inverted velocity.
gathers for every 5 km from angles 0° to 45° using the initial velocity and the inverted one are shown in Figure 14a and 14b. The angle gathers produced by the inverted velocity are flatter. The most significant aspect of the imaging result (Figure 13b) is that the pull down at a depth of 4 km under location 27.5 km is mitigated using the inverted model. Thus, the inversion captured the low-velocity anomaly at that location at depth 2.7 (Figure 9b), which is usually only captured using MVA-based tomographic methods.

**DISCUSSION**

In our implementation, the selective function is used to eliminate cycle-skipped negative contributions of the shifted correlation between the observed and modeled data. However, in some other situations, we can transform the wavefield to a pure positive function, and thus, we can compare the observed and predicted wavefields in the transformed domain. For example, we can use the square of the wavefield and obtain the following objective function:

$$\min_v \tilde{V}(r, \h)C(p^2, g^2, x, \tau, h).$$

(20)
In this case, we can use the selective function $M(\phi) = \zeta$. Otherwise, we can compare the envelope of the modeled and observed wavefields (Bozdada et al., 2011; Chi et al., 2014; Wu et al., 2014) and obtain the following objective function:

$$
\min J_2 = - \sum_h \sum_x W(x, h) C(E(p), E(g), x, h),
$$

(21)

where $E(g)$ is the envelope of the wavefield of $g$, which is defined as

$$
E(g)(t) = \sqrt{g(t)^2 + H(g)(t)^2}.
$$

(22)

In addition, $H(g)(t)$ is the Hilbert transform of $g(t)$. Even though these kinds of transformation can remove the negative contribution naturally, they lose some information about the data set in one aspect or another.

Even though the proposed method can mitigate the cycle-skipping issue for many data, it cannot fully solve the nonlinear problem in FWI, especially when the reflections dominated the data. However, we can combine this new objective function with other methods such as RWI (Wu and Alkhalifah, 2015, 2017; Alkhalifah and Wu, 2016a, 2016b) to increase the radius of convergence.

In the choice of parameters max $r$, our numerical and intuitive experience suggest that they should be large enough to encompass the shift between the modeled and observed data. As the model improves, we reduce max $r$ empirically. The process and the amount of reduction are topics for further investigation.

**CONCLUSION**

We proposed a new approach to mitigate the cycle-skipping problem. Instead of maximizing only the zero-lag correlation in the objective function, we maximize the correlation over time (or space) lag. We apply a weighting function over the lag axis that allows for a correlation of the observed and predicted data up to a user-defined maximum shift in time (or space), giving larger weights near zero lag. We use a selective function to mitigate parts of the data that contribute negatively to the objective function. With the help of the weighting function applied to the lag and a proper selective function, we obtain better convergence behavior. Numerical examples confirm these features, in which the inversion is less dependent on the initial velocity model and can tackle the cycle-skipping problem.

**ACKNOWLEDGMENTS**

We thank KAUST for its support and the SWAG group for collaborative environment. We also thank CHEVRON for providing the SEG2014 imaging challenge synthetic data set. We also thank the assistant editor J. Egen, associate editor A. Abubakar, V. Li, and two anonymous reviewers for their fruitful suggestions and comments.

**REFERENCES**


