A time-shift method for time-domain multiscale full waveform inversion
Chunhui Hu\textsuperscript{1,2}, Xiaofeng Jia\textsuperscript{1,2}, and Wenyang Chen\textsuperscript{1,2*}, 1. Laboratory of Seismology and Physics of Earth’s Interior, School of Earth and Space Sciences, University of Science and Technology of China, Anhui, China; 2. National Geophysical Observatory at Mengcheng, University of Science and Technology of China, Anhui, China.

Summary
In this study, we develop a gradient-based multiscale method in the time domain. Our time-shift method provides a natural and efficient way to realize multiscale inversion. Data preprocessing is not needed; instead, a time-shift cross-correlation is performed multiple times to extract low wavenumber gradients. Any low-pass filter applied to the gradient can be achieved by the summation of the weighted time-shift cross-correlations, in which the time shift of the wavefields is replaced with the time shift of the source wavelets. The extracted low wavenumber gradient helps reduce the dependence on the initial model in FWI.

Introduction
FWI is a strongly nonlinear optimization problem. As FWI uses a local optimization solution, the inversion process is nonunique and unstable. If the initial model is not in the vicinity of the global minimum, cycle skipping artifacts will lead the solution to converge towards a local minimum (Virieux and Operto, 2009). Multiscale inversion is a practical choice to reduce the dependence on the initial model (Bunks et al., 1995). In multiscale FWI, first the low-frequency data are inverted to recover long-wavelength velocity structures and then the high-frequency data are inverted to recover short-wavelength structures.

In the time domain, both seismic data and source wavelets have to be filtered to low-frequency bands before inversion. In recent years, the data-based multiscale methods are widely used. The most widely used filters are the Hamming window filter (Bunks et al., 1995) and Wiener filter (Boonyasiriwat et al., 2009). Ultra low-frequency data can also be obtained by frequency extrapolation (Li and Demanet, 2016), beat tone inversion (Hu, 2014), or envelope inversion (Wu et al., 2014; Liu and Zhang, 2017). Chen et al. (2015) propose a time-damping method, which uses early arrivals to invert the upper part of the model and later arrivals to invert the deeper part, thus giving a top-to-bottom solution. These methods are all data-based multiscale methods.

As recently recognized, filtering the gradient instead of the data is a promising alternative (Alkhalifah, 2015a). Tang et al. (2013) separate the gradient into tomographic and migration components, and use the tomographic component to update the long-wavelength structures. A more effective filter is based on the scattering angle and muting the low scattering angle energy of the gradient, whereas this filter has to perform 3D Fourier transform (for a 2D case) twice for each iteration (Alkhalifah, 2015a,b). Previous gradient-based methods are either inadequate or expensive.

Here we introduce a gradient-based multiscale method for the time-domain FWI, in which the gradient is low-pass filtered by a time-shift method to recover long-wavelength structures of the velocity model. We have also developed an alternative method to handle larger truncation number N to reduce the amount of computations in the case, where low-frequency data is severely over-whelmed by noise in the acquired data. We test the method using Marmousi model.

Theory
FWI is an optimization problem that minimizes the difference between the observed data and the modeled data. The difference is measured by misfit function (Tarantola, 1987)
$$ E(v) = \frac{1}{2} \|d_{\text{obs}} - d_{\text{calc}}\|_2. $$

(1)

where $d_{\text{obs}}$ denotes the observed data, $d_{\text{calc}}$ denotes the modeled data, and $\|\cdot\|_2$ is the L2 norm.

The gradient can be computed by the adjoint-gradient method (Plessix, 2006)
$$ g = \frac{\partial E(v)}{\partial v} = \frac{2}{v^2} \int_{-\infty}^{\infty} \lambda(x, t) \frac{\partial p(x,t)}{\partial z} dt, $$

(2)

According to equation (2), which has a similar form with the imaging condition, the calculation of the gradient in FWI is actually a migration process. The traditional imaging condition for prestack migration is the cross-correlation between the source wavefield and the receiver wavefield at zero time lag (Claerbout, 1971)
$$ I_0(x,z) = \sum_\tau S(x, z, \tau) R(x, z, \tau), $$

(3)

where $S$ and $R$ denote the source and receiver wavefields, respectively, and $I_0$ denotes the cross-correlation image.

The time-shift imaging condition is given by
$$ I_s(x, z, \tau) = \sum_\omega S(x, z, \omega, \tau) e^{i\omega \tau}, $$

(4)

where $I_s$ denotes the time-shift image and $\tau$ is the time shift between the source and the receiver wavefields. We can rewrite the time-shift imaging condition in the frequency domain as
$$ I_s(x, z, \tau) = \sum_\omega S(x, z, \omega) R'(x, z, \omega) e^{i\omega \tau}, $$

(5)
A time-shift method for time-domain multiscale full waveform inversion

By summing two time-shift images with opposite time shifts from each other, we can extract low-frequency contents of the image

\[ I_{low}(x, z, t) = I_l(x, z, t) + I_l(x, z, -t) \]

where \( I_{low} \) is a low-pass filtered image due to the term \( \cos(o t) \). Egen et al. (2014) uses this imaging condition to achieve a better-focused image when the migration velocity model is inaccurate. We employ this low-pass imaging condition in multiscale FWI to update long-wavelength structure. However, the direct use of this imaging condition is not effective due to the high-frequency leak of the cosine filter.

The filtering process can be expressed by

\[ I_{filt}(x, z, t) = \sum_{\omega} S(x, z, \omega) R^*(x, z, \omega) F(\omega), \]

where \( I_{filt} \) is the filtered image and \( F(\omega) \) is the low-pass filter. In Fourier cosine expansion, \( F(\omega) \) has the form of

\[ F(\omega) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos \left( \frac{n \omega}{L} \right), \]

where \( a_n \) denotes the expansion coefficients of the Fourier cosine series and \( L \) is the length of the filter \( F(\omega) \). Note that the dimension of \( L \) is the circular frequency. Thus, we have \( L = 2 \pi L_f \), in which \( L_f \) is the filter length in frequency.

Substituting equation (8) into equation (7), we have

\[ I_{filt}(x, z) = \frac{1}{2} \sum_{\omega} \sum_{n=1}^{\infty} a_n \cos \left( \frac{n \omega}{L} \right) S(x, z, \omega) R^*(x, z, \omega) \cos \left( \frac{n \omega}{L} \right) \]

Then, we have

\[ I_{filt}(x, z) = \frac{a_0}{2} I_l(x, z) + \sum_{n=1}^{\infty} a_n I_{low} \left( x, z, \frac{n \omega}{L} \right) \]

where \( a_0 \) denotes the expansion coefficients of the Fourier cosine series and \( L \) is the length of the filter \( F(\omega) \). Note that the dimension of \( L \) is the circular frequency. Thus, we have \( L = 2 \pi L_f \), in which \( L_f \) is the filter length in frequency.

Substituting equation (8) into equation (7), we have

\[ I_{filt}(x, z) = \frac{1}{2} \sum_{\omega} \sum_{n=1}^{\infty} a_n \cos \left( \frac{n \omega}{L} \right) S(x, z, \omega) R^*(x, z, \omega) \cos \left( \frac{n \omega}{L} \right) \]

Then, we have

\[ I_{filt}(x, z) = \frac{a_0}{2} I_l(x, z) + \sum_{n=1}^{\infty} a_n I_{low} \left( x, z, \frac{n \omega}{L} \right) \]

In a similar way, the filtered gradient \( g_{filt} \) can be approximated by the zero-time-lag gradient \( g_l \) and the time-shift gradients \( g_{low} \) as

\[ g_{filt} = \frac{1}{2} \left[ a_0 g_l(x, z) + \sum_{n=1}^{\infty} a_n \left[ g_l \left( x, z, \frac{n \omega}{L} \right) + g_l \left( x, z, -\frac{n \omega}{L} \right) \right] \right]. \]

In equations (10) and (11), the filtered image (gradient) can be expressed by the summation of a series of weighted time-shift cross-correlations. The summation is truncated at a flexible \( N \) for a trade-off between accuracy and computational cost. In time-shift method, we can easily extract frequency information by performing a time-shift cross-correlation multiple times without preprocessing the data. The time-shift filter is theoretically designed in the frequency domain and realized in the time domain. There is no transform switching between the time and the frequency domains to guarantee high efficiency of the method.

The filtering effect of the time-shift method is shown in Figure 1. We employ the time-shift method to approximate the target ideal filter by Fourier cosine expansion. Figure 1 shows the actual filter achieved by the time-shift method with different truncation numbers, \( N \). The optimal truncation number for Fourier cosine expansion, which makes the time-shift filter enough close to the target filter, is sought in the range of 10-20 considering both accuracy and computational cost.

The realization for large truncation number

When low frequency information is severely overwhelmed by the noise in the acquired data, a large truncation number should be adopted to construct a gradient filter that is closer to the target filter, represented by equation (11). However this requires quite a lot of cross-correlation computations. To save the computations for large truncation numbers, we develop an efficient way to calculate the filtered gradients.

Regardless of the boundary effect, equation (4) can be rewritten as

\[ I_l(x, z, t) = \sum_{\omega} S(x, z, t + \omega) R(x, z, \omega). \]

Equation 10 can be rewritten as

\[ I_{filt}(x, z) = \sum_{\omega} \sum_{n=1}^{\infty} b_n S(x, z, t + \frac{n \omega}{2L_f}). \]

Substituting equation (12) into equation (13), we have

\[ I_{filt}(x, z) = \sum_{\omega} R(x, z, t) \sum_{n=1}^{\infty} b_n S(x, z, t + \frac{n \omega}{2L_f}). \]

We now define the new source wavefield \( S'(x, z, t) \) as

\[ S'(x, z, t) = \sum_{n=1}^{\infty} b_n S(x, z, t + \frac{n \omega}{2L_f}). \]

Figure 1: The actual filter achieved by the time-shift method with different truncation numbers, \( N \).
A time-shift method for time-domain multiscale full waveform inversion

Thus, the filtered image \( I_{filt} \) can be calculated by cross-
correlating the new source wavefield \( S' \) and the receiver
wavefield \( R \). We use the operator \( T \) to denote the wavefield
propagator and \( f(t) \) to denote the source wavelet. The
original wavefield \( S \) can be expressed by
\[
S(x, z, t) = T \otimes f(t).
\]
(16)
Substituting equation (16) into equation (15), we have
\[
S'(x, z, t) = \sum_{n=-N}^{N} b_n T \otimes f \left( t + \frac{n}{2\tau_f} \right)
\]
\[
= T \otimes \sum_{n=-N}^{N} b_n f \left( t + \frac{n}{2\tau_f} \right)
\]
(17)
The new source wavefield is motivated by new modified
source wavelet, which has a relationship with the original
source wavelet as
\[
f'(t) = \sum_{n=-N}^{N} b_n f \left( t + \frac{n}{2\tau_f} \right).
\]
(18)
The filtered gradients can be achieved by the new modified
source wavelet instead of implementing the time-shift
condition multiple times. Considering the fact that we need
the original source wavelet to calculate the modeled data and
the new modified wavelet to calculate the filtered gradient,
an extra modeling process should be taken. But the large
number of cross-correlation operations in the original
method will be avoided.

Numerical examples

The 2D Marmousi model is shown in Figure 2(a). For the
consideration of the computational workload, we resampled
the Marmousi model to the size of 396×240 with the spatial
interval of 12.5 m. The data acquisition system consists of
40 shots and 396 receivers evenly distributed along the
surface. The observed data is generated using a 10-Hz Ricker
wavelet, with a record length of 4 s sampled at 1 ms. The
initial model is shown in Figure 2(b), which is a highly
smoothed version of the true model.

The time-shift multiscale method is used to build the
background model, which is then used as the initial model
for traditional FWI to recover the small-scale details.

The model is inverted with two scales in frequency: 0-8 Hz
and full-band frequencies. The length of the target filter is 25
Hz; thus, the unit time shift is \( \tau = \frac{1}{2\tau_f} = 0.02 \) s. In Fourier
cosine expansion, the target filter is expanded to 12 cosine
terms, which means we need to implement the time-shift
cross-correlation 24 times according to equation (11). The
weighted coefficients \( a_n \) are shown in Table 1.

Figure 3 shows the FWI gradient and the time-shift filtered
gradient at the first iteration. The gradient with the time-shift
method is much smoother than the original one and is used
successfully in building the long-wavelength structures. In
Figure 4(a), the background velocity model is achieved by
the time-shift FWI with 50 iterations. The background model
is then used as the initial model for subsequent traditional
FWI. After another 100 iterations, the final result is obtained
in Figure 4(b). For comparison, the single-scale FWI result
is shown in Figure 4(c). Figure 4(b) provides more accurate
details than Figure 4(c), especially in the anticline and fault
areas of the central model.

Table 1: Coefficients \( a_n \) for the Marmousi test.

<table>
<thead>
<tr>
<th>Time shift</th>
<th>0</th>
<th>±( \tau )</th>
<th>±2( \tau )</th>
<th>±3( \tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients ( a_n )</td>
<td>0.7200</td>
<td>0.5375</td>
<td>0.2880</td>
<td>0.0266</td>
</tr>
<tr>
<td>Time shift</td>
<td>...</td>
<td>±10( \tau )</td>
<td>±11( \tau )</td>
<td>±12( \tau )</td>
</tr>
<tr>
<td>Coefficients ( a_n )</td>
<td>...</td>
<td>-0.0374</td>
<td>-0.0578</td>
<td>-0.0256</td>
</tr>
</tbody>
</table>
A time-shift method for time-domain multiscale full waveform inversion

In the above numerical test, we should implement the time-shift cross-correlation 24 times in each iteration according to equation (11). To avoid large amounts of cross-correlation, we use the modified source wavelet given by equation (18) to calculate the filtered gradients, instead of implanting the time-shift condition multiple times. The original 10-Hz Ricker wavelet and the new modified source wavelet are shown in Figure 5.

In Figure 6, the background velocity model is shown, which is achieved by using the modified source wavelet instead of the original Ricker wavelet with 50 iterations. The background velocity model in Figure 6 is not much different from the background velocity model in Figure 4(a). By using the modified source wavelet, one more forward calculation will be performed for each iteration, but the large number of cross-correlation operations in the original method will be avoided. In the numerical example of this abstract, one third of the calculation time was reduced compared with the original method. It is clear that the efficiency saving increases as the number of cross-correlation increases, which should be determined by the needs of specific inversion problem.

Conclusion

Unlike commonly used data-based multiscale methods, the time-shift multiscale method is a gradient-based method in the time domain. Low-wavenumber components of the gradient can be extracted using the time-shift method. The extraction is realized by performing the time-shift cross-correlation multiple times instead of low-pass filtering of the seismic data. We use the time-shift method to recover the long-wavelength structures of the model and use traditional FWI to recover the short-wavelength structures, thus implementing a multiscale inversion scheme. The time-shift method helps reduce the dependence of the initial model and avoid local minima. The convergence rate is accelerated. We have also developed an alternative method to handle larger truncation number to avoid the large amount of cross-correlation calculations used to deal with the case, where low-frequency data is severely over-whelmed by noise in the acquired data.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (41774121) and the Fundamental Research Funds for the Central Universities (WK2080000078).
REFERENCES


