

# Frequency domain adaptive waveform inversion

Matt Eaid, Scott Keating and Kris Innanen

## ABSTRACT

Full waveform inversion (FWI) attempts to find a high-resolution model of subsurface parameters, carrying a high likelihood of having produced the observed seismic data. While classic FWI is generally quite successful, the high degree of nonlinearity involved in seismic inverse problems, coupled with the oscillatory nature of seismic data can invoke a phenomenon known as cycle skipping, leading to locally minimized objective functions. Generally, when cycle skipping occurs the updated model is a worse representation of the true subsurface than the starting model was.

Extended waveform inversion is a relatively new idea that is an umbrella for a suite of inversion techniques that extend the model space, usually by some nonphysical parameter, and then drive that parameter to an ideal quantity, matching the predicted to the observed data. They combat the cycle skipping problem by adding a degree of freedom to the model space and forming objective functions that do not rely on sample-by-sample differences. The flavour of extended waveform inversion we present is known as adaptive waveform inversion (AWI) which extends the model space in convolutional Wiener filter coefficients and attempts to drive them towards a zero-lag delta spike. Originally derived in the time domain, we present a frequency domain alternative and discuss special considerations for frequency domain implementation. We then show one example where AWI is more robust than FWI and discuss the challenges going forward.

## INTRODUCTION

Full waveform inversion obtains high resolution models of subsurface parameters by minimizing an objective function that relies on some comparison of the observed and predicted wavefields (Lailly, 1983; Tarantola, 1984). Predicted data is generated by propagating wavefields in a subsurface model, and then compared to the observed field data according to some measure. The model is then updated in an iterative fashion until the predicted and observed wavefields are in sufficient agreement, according to this measure, producing a model that has a high likelihood of having produced the field data. The most common form of the FWI objective function, known as least-squares FWI (LS-FWI), compares the sample-by-sample difference between the predicted and observed wavefields, known as residuals, and attempts to minimize the sum of the squared residuals (Virieux and Operto, 2009). While this method works well most of the time, the oscillatory nature of seismic data creates challenges for sample-by-sample comparisons. If the starting model and true model lead to predicted and observed data that are out of phase by more than a half cycle, then the least-squares objective function can be locally minimized by updating the model such that neighboring cycles of the observed and predicted data are aligned; a problem known as cycle skipping. This is an obvious problem, in which the FWI algorithm gets trapped in a local minimum and fails to update towards the true model, producing an updated model that is in worse agreement with the true model than the starting model was.

Fortunately, this is not the only way to formulate a successful FWI objective function.

Recently, many publications have attempted to mitigate the effects of cycle skipping by extending the model space in one of several candidate nonphysical parameters (Symes, 2008; van Leeuwen and Mulder, 2010; Luo and Sava, 2011), allowing for the development of objective functions that limit local minima, by altering the way in which the predicted data is matched to the observed data. One such method, known as adaptive waveform inversion (Warner and Guasch, 2016) extends the model space in convolutional Wiener filter coefficients that when convolved with the predicted data provide a least squares match to the observed data. It then attempts to drive these coefficients towards those of an identity filter that leaves its input unchanged. As we will see this procedure drives the ratio between the observed and predicted data to unity, by solving an objective function that is more robust to problems associated with cycle skipping. Warner and Guasch (2016) presented their derivation and results based on a time-domain formulation of the full waveform inversion method. Given the growing interest in frequency domain methods we derive a frequency domain version of the adaptive waveform inversion procedure based on the work of Warner and Guasch (2016). We then show examples of how it better handles models that lead to cycle skipped data as compared to LS-FWI

## THEORY

### Least-squares full waveform inversion

The original formulation of FWI attempts to minimize the sample-by-sample difference between predicted data generated from a postulated model and the observed data recovered from the true model by updating the model until the datasets agree. The goal is to generate a high resolution model of the subsurface parameters that most likely generated the observed data by minimizing an objective function of the form,

$$f = \frac{1}{2} \sum_{r_g, r_s, \omega} \delta \mathbf{d}^T \delta \mathbf{d}^* \quad (1)$$

where  $\delta \mathbf{d}$  is the residual given by  $\delta \mathbf{d}_i = \mathbf{u}_i - \mathbf{d}_i$ , with the predicted data given by  $\mathbf{u}$  and the observed data given by  $\mathbf{d}$ .

Minimization of equation (1) is achieved by taking the derivative with respect to the model ( $\mathbf{m}$ ) and setting it equal to zero.

$$\frac{\partial f}{\partial \mathbf{m}(\mathbf{r})} = \sum_{r_g, r_s, \omega} \frac{\partial \mathbf{u}^T}{\partial \mathbf{m}(\mathbf{r})} \delta \mathbf{d}^* \quad (2)$$

where the quantity  $\partial \mathbf{u}^T / \partial \mathbf{m}(\mathbf{r})$  in equation (2) is known as the sensitivity. To determine the sensitivity for our given forward problem we may start from any general wave equation, given in operator notation by equation (3).

$$\mathcal{L}(\mathbf{r}_g, \omega | \mathbf{m}) \mathbf{G}(\mathbf{r}_g, \mathbf{r}_s, \omega | \mathbf{m}) = \delta(\mathbf{r}_g - \mathbf{r}_s) \quad (3)$$

Taking the derivative of equation (3) with respect to the model and setting it equal to zero, gives us the sensitivity,

$$\frac{\partial \mathbf{G}(\mathbf{r}_g, \mathbf{r}_s, \omega | \mathbf{m})}{\partial \mathbf{m}(\mathbf{r})} = -\mathcal{L}^{-1}(\mathbf{r}_g, \omega | \mathbf{m}) \frac{\partial \mathcal{L}(\mathbf{r}_g, \omega | \mathbf{m})}{\partial \mathbf{m}(\mathbf{r})} \mathbf{G}(\mathbf{r}_g, \mathbf{r}_s, \omega | \mathbf{m}) \quad (4)$$

In the acoustic case the wave operator given by  $\mathcal{L}$  in equation (4) is  $\mathcal{L} = \nabla^2 + \omega^2 \mathbf{m}(\mathbf{r}_g)$  where  $\mathbf{m}(\mathbf{r}_g)$  is the slowness. Substituting this into equation (4) and using the theory of Green's function the acoustic wave equation sensitivity is given by equation (5).

$$\frac{\partial \mathbf{G}(\mathbf{r}_g, \mathbf{r}_s, \omega | \mathbf{m})}{\partial \mathbf{m}(\mathbf{r})} = -\omega^2 \mathbf{G}(\mathbf{r}, \mathbf{r}_s, \omega | \mathbf{m}) \mathbf{G}(\mathbf{r}_g, \mathbf{r}, \omega | \mathbf{m}) \quad (5)$$

Substituting the sensitivity given by equation (5) into equation (2) gives us the gradient required to minimize the least squares objective function

$$\frac{\partial f}{\partial \mathbf{m}(\mathbf{r})} = \mathbf{g}(\mathbf{r}) = - \sum_{r_g, r_s, \omega} \omega^2 \mathbf{G}(\mathbf{r}, \mathbf{r}_s, \omega | \mathbf{m}) \mathbf{G}(\mathbf{r}_g, \mathbf{r}, \omega | \mathbf{m}) \delta \mathbf{d}^*(\mathbf{r}_g, \mathbf{r}_s, \omega | \mathbf{m}) \quad (6)$$

Equation (5), read from right to left says that we first compute the residual, then reverse it in time and propagate it from the receiver to a point  $\mathbf{r}$ . This receiver wavefield is then correlated with the forward propagated wavefield and weighted by the frequency, forming the gradient. The model is then updated by  $\mathbf{m}_i = \mathbf{m}_0 + \alpha \mathbf{g}(\mathbf{r})$ , where  $\alpha$  is a step length. As discussed earlier this objective function contains many local minima, and the gradient can update the model to one which is actually further from the truth than the starting model. The next section will examine the properties of this objective function that make it prone to errors when the data are cycle skipped and will provide insight into how to develop objective functions that are more immune to cycle skipping.

### Cycle skipping and the problem of local minima

Consider the case of figure 1(a) in which we have observed data from the true model in black, and data from a proposed starting model in red. In this case our starting model contains sufficient error, such that the predicted data is out of phase with the observed data by more than a half wavelength. A desirable FWI formulation would shift the predicted data to the left, matching the observed data as shown in figure 1(b). However, due to the fact that least-squares FWI minimizes the sample-by-sample difference between the observed and predicted data, the predicted data would erroneously be shifted to the right (figure 1(c)). The reason for this, is on a sample-by-sample basis the only way the LS-FWI objective can be minimized is by matching the first trough of the predicted data to the second trough of the observed data. Thus we have matched adjacent cycles of the observed and predicted data, and we may say the data are cycle skipped. In figure 1(c) we have found a local minimum of the LS-FWI objective function and in this case the inversion would no longer update the model, we are essentially stuck in this local minimum.

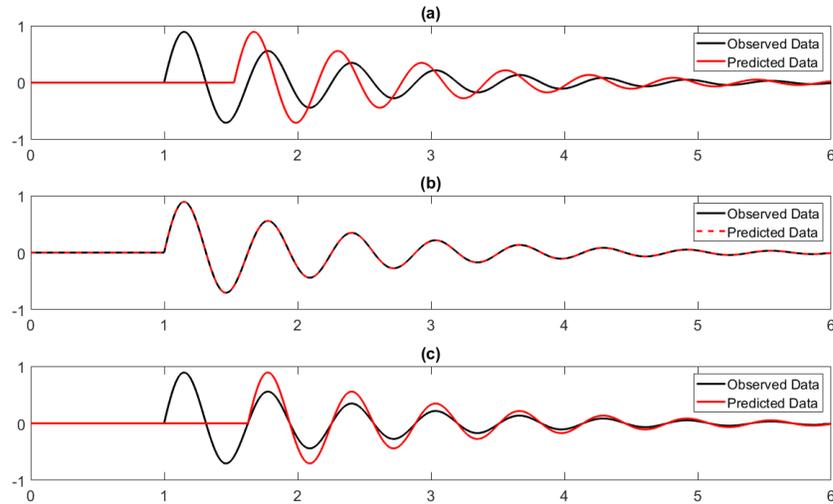


FIG. 1. (a) Observed data (black), data from a model that would induce cycle skipping (red); (b) Data from a model that has been updated to match the true model (red); (c) Same observed data from (a) in black, data resulting from the model recovered by least-squares FWI (red), leading to cycle skipping.

### Adaptive waveform inversion

Conventional formulations of full waveform inversion, while successful, suffer when the starting model is sufficiently erroneous to induce cycle skipping. This is due both to the oscillatory nature of seismic data, and the local minima contained in the LS-FWI objective function. Other successful formulations of the FWI objective function exist, and are more robust to cycle skipping problems. Most of these formulations extend the model space in some nonphysical parameter to help develop a more desirable objective function. One such extension shown by Warner and Guasch (2016), expands the model space in convolutional Wiener filter coefficients, and is known as adaptive waveform inversion (AWI).

Adaptive waveform inversion computes a matching filter  $\mathbf{w}$  that when convolved with the predicted data trace  $\mathbf{p}$  provides the least-squared fit to the observed data trace  $\mathbf{d}$ . The filter ( $\mathbf{w}$ ) is computed through solving equation (7),

$$\mathbf{g} = \frac{1}{2} \|\mathbf{P}\mathbf{w} - \mathbf{d}\|^2, \quad (7)$$

where  $\mathbf{P}$  is a Toeplitz matrix representing convolution of the predicted data with the filter  $\mathbf{w}$ . Equation (7) has the well known least-squares solution,

$$\mathbf{w} = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{d}. \quad (8)$$

AWI then seeks an identity filter, which passes through its input unchanged. For convolutional Wiener filters the identity filter is a delta spike at zero lag. Because the filter

depends on both the observed and predicted data, which necessarily depend on the true and predicted model, driving this filter to a zero lag delta spike can only be achieved by driving the predicted model to the true model, such that the predicted and observed data match. Keeping this in mind, Warner and Guasch (2016) choose an objective function that when minimized, forces  $\mathbf{w}$  towards a zero lag delta spike as shown in equation 9.

$$f = \frac{1}{2} \frac{\|\mathbf{T}\mathbf{w}\|^2}{\|\mathbf{w}\|^2} \quad (9)$$

In equation (9)  $\mathbf{T}$  is a diagonal matrix that helps drive the filter to a zero lag delta spike by weighting the coefficients of  $\mathbf{w}$  by a monotonically increasing function of their temporal lag.

*Adaptive waveform inversion: gradient and adjoint source*

Adaptive waveform inversion, similarly to FWI proceeds by finding the sensitivity  $\partial\mathbf{p}/\partial\mathbf{m}$  of a given wave equation  $\mathbf{A}\mathbf{u} = \mathbf{S}$ ,

$$\frac{\partial\mathbf{p}}{\partial\mathbf{m}} = -\mathbf{R}\mathbf{A}^{-1} \frac{\partial\mathbf{A}}{\partial\mathbf{m}} \mathbf{u} \quad (10)$$

where  $\mathbf{p} = \mathbf{R}\mathbf{u}$  is the wavefield extracted at the receiver coordinates given by  $\mathbf{R}$ . Following from Pratt et al. (1998) the gradient of any objective function of the form  $f = (1/2)\mathbf{x}^T\mathbf{x}$  is given by equation (11),

$$\frac{\partial f}{\partial\mathbf{m}} = -\mathbf{u}^T \left( \frac{\partial\mathbf{A}}{\partial\mathbf{m}} \right)^T \mathbf{A}^T \mathbf{R}^T \frac{\partial\mathbf{x}^T}{\partial\mathbf{p}} \mathbf{x} = -\mathbf{u}^T \left( \frac{\partial\mathbf{A}}{\partial\mathbf{m}} \right)^T \mathbf{A}^T \mathbf{R}^T \delta\mathbf{s}. \quad (11)$$

The term  $\delta\mathbf{s}$  is the adjoint source used for back propagation of the receiver wavefield, which for LS-FWI was given by the residuals.

Provided FWI and AWI assume the same wave physics, the only term in equation (11) that differs between the two formulations is the adjoint source. Equation (11) shows that the adjoint source is given by the derivative of the chosen objective function with respect to the predicted data,

$$\delta\mathbf{s}_{\text{AWI}} = \frac{\partial f}{\partial\mathbf{p}} = \frac{1}{2} \frac{\partial}{\partial\mathbf{p}} \frac{\mathbf{w}^T \mathbf{T}^2 \mathbf{w}}{\mathbf{w}^T \mathbf{w}} = \mathbf{W}^T \mathbf{P} (\mathbf{P}^T \mathbf{P})^{-1} \left( \frac{2f\mathbf{I} - \mathbf{T}^2}{\mathbf{w}^T \mathbf{w}} \right) \mathbf{w} \quad (12)$$

where,

$$\frac{\partial\mathbf{w}}{\partial\mathbf{p}} = (\mathbf{P}^T \mathbf{P})^{-1} \frac{\partial\mathbf{P}^T}{\partial\mathbf{p}} \mathbf{d} + \frac{\partial(\mathbf{P}^T \mathbf{P})^{-1}}{\partial\mathbf{p}} \mathbf{P}^T \mathbf{d} = -(\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \mathbf{W}. \quad (13)$$

The adjoint in AWI is somewhat more complicated than the data residual employed in LSFWI, we must first find the convolutional Wiener filter ( $\mathbf{w}$ ) that matches the predicted to the observed data. We then normalize this filter by its  $L_2$  norm ( $\mathbf{w}^T \mathbf{w}$ ), and weight it by coefficients that depend on the form of  $\mathbf{T}$ , and our given objective function ( $2f\mathbf{I} - \mathbf{T}^2$ ). This result is then deconvolved by the auto-correlation of the predicted trace, given by  $(\mathbf{P}^T \mathbf{P})^{-1}$ , which is then convolved with the predicted trace, and correlated with the filter  $\mathbf{W}$ . This adjoint differentiates AWI from FWI and is at the heart of the AWI method. The key concept of AWI is that its objective and therefore adjoint no longer compare oscillatory data sets sample-by-sample which manifests in a more complex adjoint source that is more robust in the presence of cycle skipped data.

### *Frequency domain adaptive waveform inversion*

Adaptive waveform inversion, originally derived in the time domain (Warner and Guasch, 2016), solves for the Wiener filter coefficients through time domain convolution using Toeplitz matrices. In the original formulation Wiener filters are computed for every source-receiver pair, which is fairly efficient on a trace-by-trace basis, but becomes computationally cumbersome when the dataset is composed of thousands of traces. This, along with general interest in frequency domain methods, motivates the interest in a frequency domain formulation of AWI.

Warner and Guasch (2016) experimented with spatial Wiener filters, one filter for all source-receiver pairs at a given time sample, but observed poor convergence. However, they discuss that these spatial filters may see more success in a frequency domain formulation, where we would have one filter for all source-receiver pairs at a given frequency. This is beneficial in FWI, where we typically have far fewer frequencies of interest than source receiver pairs, greatly reducing the number of filter computations required. Additionally, the frequency domain filters, and thus the adjoint will rely on multiplications instead of convolutions in the frequency domain, further reducing run time. Since convolution in the time domain is frequency domain multiplication, equation (8) becomes,

$$\mathbf{w} = \frac{\mathbf{p}^* \mathbf{d}}{\mathbf{p}^* \mathbf{p} + \mu A_{\text{MAX}}} \quad (14)$$

where  $\mathbf{p}^*$  represents complex conjugation,  $\mu$  is a stability term, and  $A_{\text{MAX}}$  is the maximum value of the amplitude spectrum of the predicted trace.

Computing the adjoint progresses in a similar fashion to the time domain formulation, giving the frequency domain adjoint,

$$\delta_s = \frac{\mathbf{w}^* \mathbf{p}}{\mathbf{p}^* \mathbf{p}} \left( \frac{2f\mathbf{I} - \mathbf{T}^2}{\mathbf{w}^\dagger \mathbf{w}} \right) \mathbf{w} \quad (15)$$

where  $\mathbf{w}^\dagger$  is the complex conjugate transpose. Recall that the identity filter sought in the frequency domain, is equal to one at all frequencies. Studying equation (14), this occurs

when  $\mathbf{p} = \mathbf{d}$ , or when the predicted data is equal to the observed data. AWI and FWI both update the model to improve the agreement between the predicted and observed data. FWI achieves this by driving the difference between the observed and predicted data to zero, while AWI approaches the problem by driving their ratio to unity,

$$\text{FWI} : \mathbf{p} - \mathbf{d} \rightarrow \mathbf{0} \qquad \text{AWI} : \frac{\mathbf{p}}{\mathbf{d}} \rightarrow \mathbf{1}. \qquad (16)$$

*Frequency domain weighting matrix*

In the time domain, the weighting matrix  $\mathbf{T}$  acts to penalize filter coefficients at large temporal lag, forcing the filter towards a zero lag spike as the algorithm iterates. Due to the intimate relationship between time and phase, it might be tempting in the frequency domain to design the weighting matrix such that it penalizes some aspect of the phase. However, there is no physical aspect of the phase that is equivalent to large temporal lags, making design of a weighting matrix based on phase challenging. We instead take a different approach, by keeping in mind the filter we are after is a zero lag delta spike, or in frequency, a filter of ones at all frequencies. We could then formulate  $\mathbf{T}$  such that our objective minimizes the difference between our filter and a vector of ones.

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & -1 \\ 0 & 1 & 0 & 0 & \dots & -1 \\ 0 & 0 & 1 & 0 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \dots & -1 \\ 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ 1 \end{bmatrix} \qquad (17)$$

Using the special formulation of  $\mathbf{T}$  given by equation (17), the objective function of equation (9) can be rewritten as,

$$f = \|\mathbf{w} - \vec{1}\|^2 \qquad (18)$$

which now directly minimizes the difference between the computed filter and the identity filter. However, this formulation is not particularly robust to cycle skipping, due to the fact that this formulation does not weight individual components of the filter.

We briefly return discussion to the time domain for ease of visualization. The filter sought, in the time domain, is a zero lag delta spike. Figure 2 shows how the amount of lag is related to frequency. Looking at the upper left portion of figure 2, filters with large lag have a low frequency component when compared to the identity filter. In contrast, filters with small lags contain an inherent short wavelength component. This is further emphasized on the bottom of figure 2 which shows the envelope of the amplitude spectrum for the large lag filter in blue, and the small lag filter in red. In wanting to penalize coefficients at large temporal lags to drive our filter towards an identity filter, it is evident that in the frequency domain we could penalize errors in our filter at low frequencies. To accomplish

this we may multiply the form  $\mathbf{T}$  takes in equation (17) by any function which is large for low frequencies and small for large frequencies.

$$\mathbf{T} = \frac{1}{\sqrt{2\pi\sigma_f^2}} \exp\left(\frac{-\omega^2}{2\sigma_f^2}\right) \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & -1 \\ 0 & 1 & 0 & 0 & \dots & -1 \\ 0 & 0 & 1 & 0 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \dots & -1 \\ 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix} \quad (19)$$

In the following examples a Gaussian of the form in equation (19) is employed, where  $\sigma_f$  should be chosen to sufficiently penalize low frequencies.

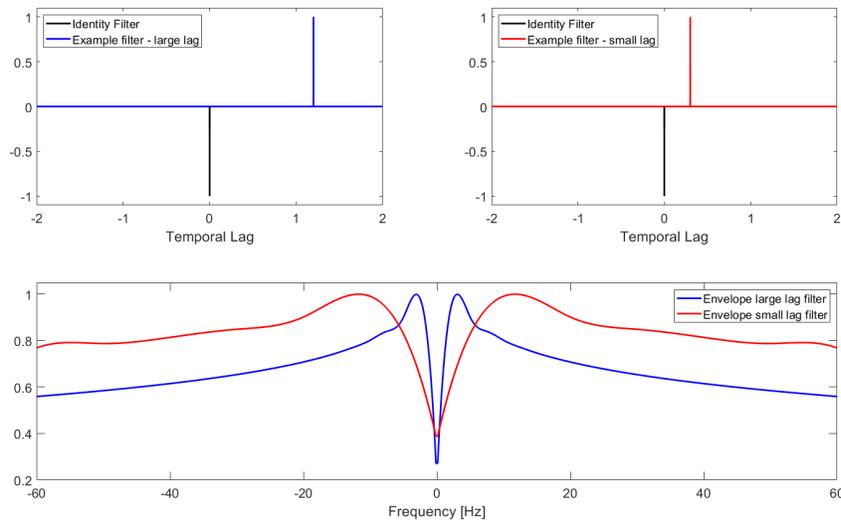


FIG. 2. (Top left) Example of large lag filter (blue) and identity filter (black), (top right) example of small lag filter (red) and identity filter (black), (bottom) envelope of amplitude spectrum formed by subtraction of the identity filter from the example filter for the cases in the top two figures.

## EXAMPLES

### Example 1: Ball Model

The first example is one in which cycle skipping is not expected to take place so that we may test the efficacy of adaptive waveform inversion as compared to full waveform inversion. The true velocity model contains two ball shaped anomalies, one of lower velocity than the background medium, and one of higher velocity (figure 3). Sources and receivers were placed in a vertical line at a distance of 20 meters from the left edge of the model. The source spacing was set at 20 meters, while the receivers were placed at 10 meter intervals. A second vertical line of sources and receivers was placed 490 meters from the left edge of the model, simulating a cross well experiment. The starting model was chosen to be constant with a velocity equal to the background velocity of the true model (2500 m/s).

Figure 4(a) shows the inversion result obtained by frequency domain full waveform

inversion, while figure 4(b) shows the result obtained through frequency domain adaptive waveform inversion. The respective differences between the true and inverted models are shown in the bottom row. Both inversions were carried out using a multi-scale (Bunks, 1995) approach with frequencies from 3-20 Hz, using 5 bands, with 10 iterations per band. Although subtle differences exist between the two inversion results, FWI and AWI arrive at a comparable result, showing that AWI is comparable to FWI in this case.

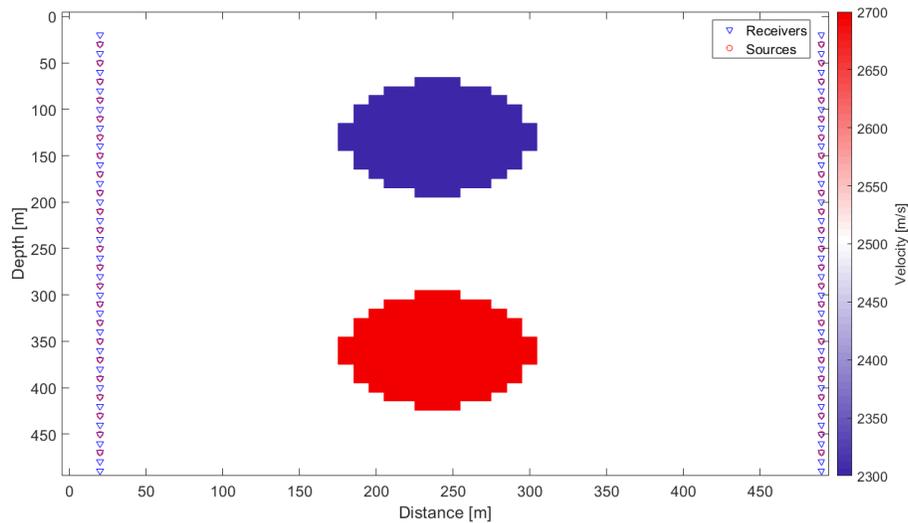


FIG. 3. True velocity model producing data that is not cycle skipped, with acquisition geometry shown, receivers are marked by blue triangles, and sources by red circles.

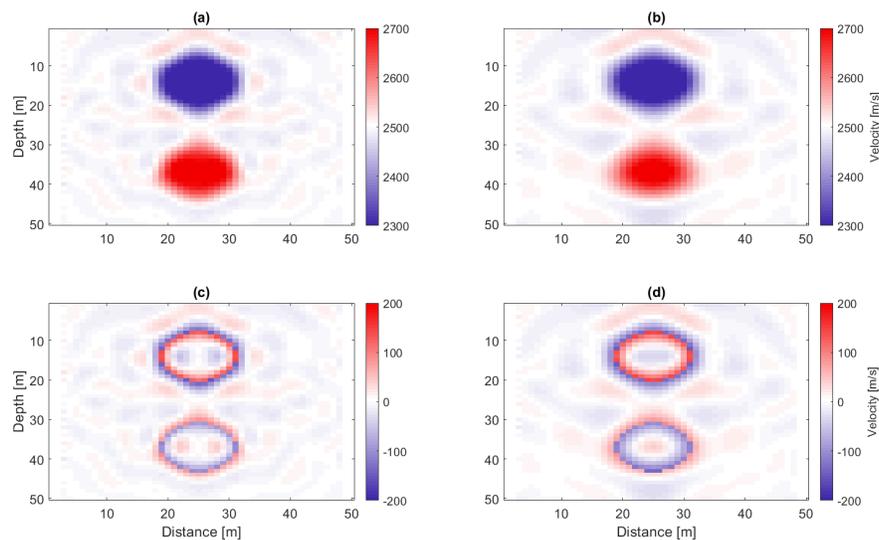


FIG. 4. Inversion results from model in figure 3. (a) Velocity model inverted by frequency domain FWI, (b) Velocity model inverted by frequency domain AWI, (c) Difference between the velocity model in (a) and the true velocity model, (d) Difference between the velocity model in (d) and the true velocity model.

## Example 2: Gaussian Anomaly

The next example is one in which cycle skipping occurs, and has damaging effects on the full waveform inversion algorithm. Figure 5 shows the true velocity model which contains a Gaussian anomaly embedded in a constant background velocity of 2000 m/s. Figure 5 also shows the acquisition geometry which contains a vertical line of sources and receivers placed 40 meters from the left edge of the model, and a vertical line of sources and receivers at 2000 meters from the left edge of the model, again simulating a cross well experiment. The source spacing was set at 40 meters, and the receiver spacing at 20 meters.

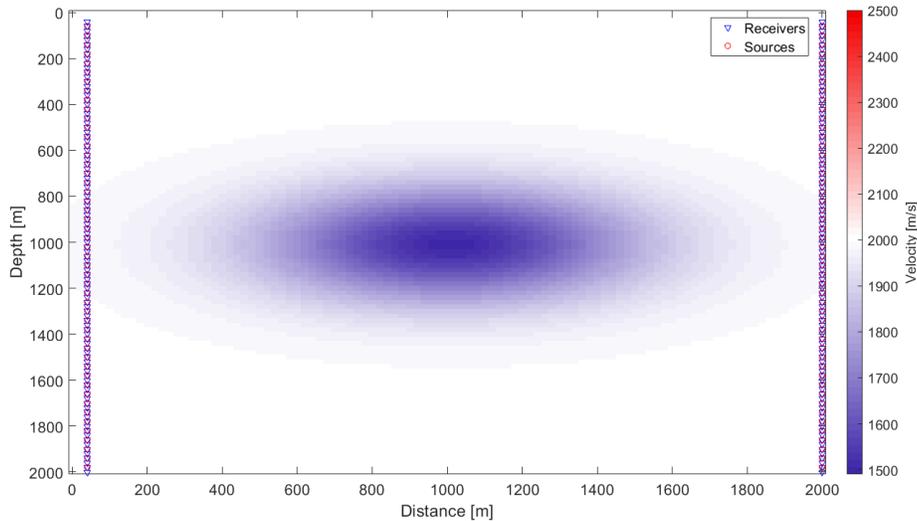


FIG. 5. True velocity model producing data that is cycle skipped, with acquisition geometry shown, receivers are marked by blue triangles, and sources by red circles.

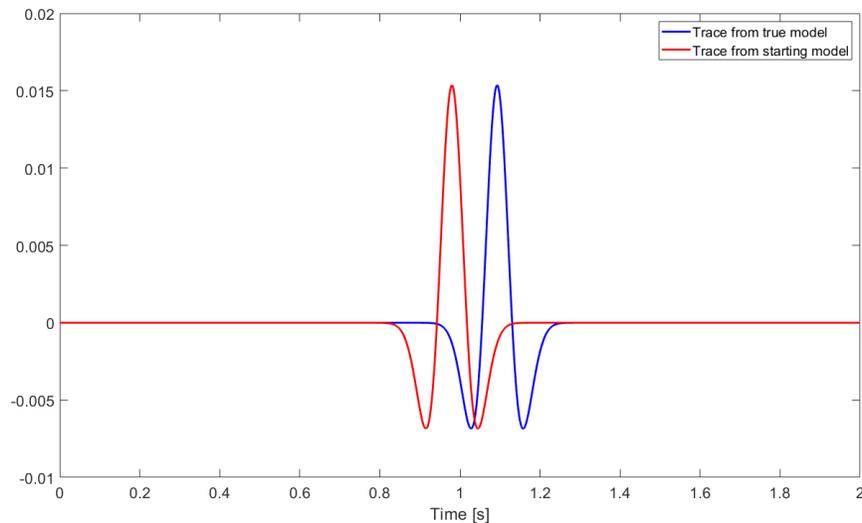


FIG. 6. Traces from source and receiver at a depth of 1000 meters, for the starting model (red) and true model of figure 5 (blue)

Figure 6 shows traces obtained from the source-receiver pair at 1000 meters depth for

the starting model, which is constant (red), and for the true model of figure 5 (blue). This example highlights a clear case of cycle skipped data, a successful inversion algorithm should decrease the starting velocity model so that the peak of the red trace matches the peak of the blue trace. Unfortunately a least-squares FWI formulation will update the velocity model such that the second trough of the red trace matches the first trough of the blue trace, which requires a velocity increase. This represents a dire case where FWI updates the starting model so that it is a worse match to the true model.

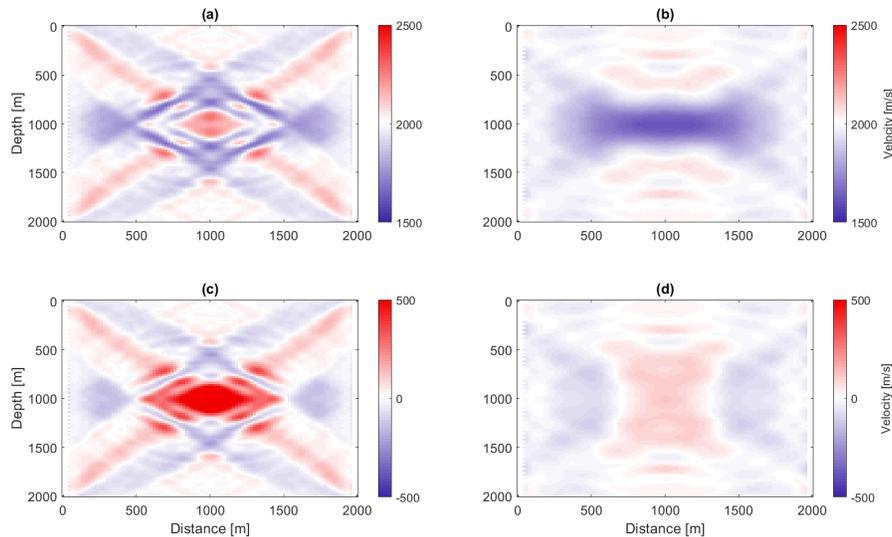


FIG. 7. Inversion results from model in figure 5.(a) Velocity model inverted by frequency domain FWI, (b) Velocity model inverted by frequency domain AWI,(c) Difference between the velocity model in (a) and the true velocity model, (d) Difference between the velocity model in (d) and the true velocity model.

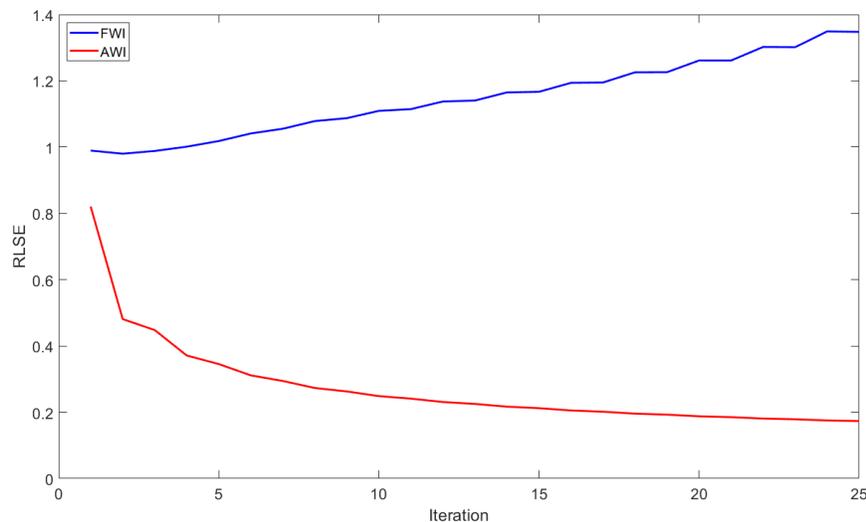


FIG. 8. Relative least-squared model error versus iteration number for FWI in blue and AWI in red, for the Gaussian model.

Figure 7 shows the velocity model inverted by FWI in (a) and the velocity model inverted by AWI in (b) with the respective differences shown in the bottom row. Figure 7(a)

shows a clear case of FWI producing a very poor result in the presence of cycle skipped data. FWI has significantly increased the velocity in the center of the model, where it should have been decreased. The difference in figure 7(c) shows poor agreement between the true and inverted models. Conversely the inverted model obtained through AWI (figure 7(b)) is a very good match to the true starting model, which is highlighted by the weak magnitudes contained in the plot of the difference between the true and inverted models (figure 7(d)). Figure 8 shows the relative least-squares error (RLSE)  $\epsilon$ , which is a measure of the proximity of the inverted model to the true model relative to the starting model, for FWI in blue and AWI in red. Where  $\epsilon$  is given by,

$$\epsilon = \frac{\|\mathbf{m}_k - \mathbf{m}_{\text{TRUE}}\|^2}{\|\mathbf{m}_0 - \mathbf{m}_{\text{TRUE}}\|^2}. \quad (20)$$

If our FWI implementation performs desirably, that is it obtains a good estimate of the subsurface properties, then  $\epsilon$  will be driven to lower values as the iterations progress. Figure 8 emphasizes that the AWI algorithm, shown in red, is performing desirably. However, as the iterations proceed it is evident that FWI shown by the blue line is giving us a worse estimate of the true model than we initially had.

## CONCLUSIONS

Cycle skipping is a problem that can hinder our ability to recover accurate subsurface models, arising from the way in which we choose to formulate our objective function. The conventional formulation of FWI in which we try to, in a least-squares sense, minimize the sample-by-sample difference between the observed and predicted data is especially prone to cycle skipped data. For reasons outlined above, the nonlinear nature of seismic inverse problems coupled with oscillatory nature of seismic problems results in LS-FWI formulations getting trapped in local minima or producing poorly inverted models in the presence of cycle skipped data. One way to mitigate the effects of cycle skipping is to extend the model space in some nonphysical parameter, allowing for reformulation of the objective function in such a way that we are matching predicted to observed data without comparing them sample-by-sample. Adaptive waveform inversion achieves this by extending the model space in Wiener filter coefficients, allowing us to formulate the objective so that the ratio between the observed and predicted data is driven to unity.

Adaptive waveform inversion was originally derived in the time domain. Here, we show a reformulation in the frequency domain, which reduces the computational burden associated with convolutions, and allows us to apply one filter to all source-receiver pairs instead of one filter for every source-receiver pair. Additionally, it allows us to employ AWI in the increasingly popular frequency domain representation of waveform inversion. We then discussed strategies for developing the weighting matrix in the frequency domain and discussed its ties to the time domain weighting matrix. The frequency domain formulation of AWI was then tested on two models, a ball model that did not produce cycle skipped data, and a Gaussian anomaly that induced cycle skipping. The first model was used to compare the AWI formulation to the FWI formulation in a case where both methods are expected to perform well. Apart from subtle differences both methods produced compatible results.

The second model, consisting of the Gaussian anomaly, produced data that was severely cycle skipped when compared to the data from the starting model. Adaptive waveform inversion was able to successfully recover a good estimate of the true model, while full waveform inversion resulted in a model that was grossly in error when compared to the true model. These results confirm that our frequency domain formulation is a successful approach to mitigate cycle skipping.

## FUTURE WORK

All of the current work in adaptive waveform inversion has focused purely on gradient-based steepest descent optimization. It has been well documented that in many cases more complicated Newton type steps are required to mitigate problems associated with FWI, as well as provide faster convergence. The next logical step, now that we have a gradient formulated for AWI is to develop the terms required to compute the Hessian and study AWI in Newton type steps. Additionally, we would like to test the AWI scheme under more sophisticated optimization algorithms and compare its convergence rate to FWI. One of the main flaws with our formulation is that it is not as robust as the time domain version, especially in the presence of missing low frequency information. The choice of  $\mathbf{T}$  is important to combat cycle skipping, devising a more robust formulation for  $\mathbf{T}$  is a concern and will be the focus of continued research.

## ACKNOWLEDGMENTS

The authors would like to thank the sponsors of the CREWES project as well NSERC under the grant CRDPJ 461179-13 for making this work possible through their financial support.

## REFERENCES

- Bunks, C., 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**.
- Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, Society for Industrial and Applied Mathematics, Expanded Abstracts.
- Luo, S., and Sava, P., 2011, A deconvolution-based objective function for wave-equation inversion: SEG Expanded Abstracts 81 International Meeting.
- Pratt, R., Shin, C., and Hicks, G., 1998, Gauss-newton and full newton methods in frequency-space seismic waveform inversion: *Geophysics Journal International*.
- Symes, W., 2008, Migration velocity analysis and waveform inversion: *Geophysical Prospecting*, , No. 6.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*.
- van Leeuwen, T., and Mulder, W., 2010, A correlation-based misfit criterion for wave-equation travel time tomography: *Geophysics Journal International*.
- Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**.
- Warner, M., and Guasch, L., 2016, Adaptive waveform inversion: theory: *Geophysics*, , No. 6.