We E106 13

# Adaptive Waveform Inversion - FWI Without Cycle Skipping - Theory

M. Warner* (Imperial College London) & L. Guasch (Sub Salt Solutions Ltd)

## SUMMARY

Conventional FWI minimises the direct differences between observed and predicted seismic datasets. Because seismic data are oscillatory, this approach will suffer from the detrimental effects of cycle skipping if the starting model is inaccurate. We reformulate FWI so that it instead adapts the predicted data to the observed data using Wiener filters, and then iterates to improve the model by forcing the Wiener filters towards zero-lag delta functions. This adaptive FWI scheme is demonstrated on synthetic data where it is shown to be immune to cycle skipping, and is able to invert successfully data for which conventional FWI fails entirely. The new method does not require low frequencies or a highly accurate starting model to be successful. Adaptive FWI has some features in common with wave-equation migration velocity analysis, but it works for all types of arrivals including multiples and refractions, and it does not have the high computational costs of WEMVA in 3D.

## Introduction

Full-waveform inversion (FWI), is a technology that able, under appropriate circumstances, to generate high-resolution high-fidelity models of seismic velocity and other physical properties in the subsurface. These models can be used both to improve the subsequent migration of conventional reflection data, and to provide direct information about the reservoir and its environs. FWI attempts to find a model of the subsurface that predicts the entire recorded seismic wavefield, and it typically proceeds via a series of linearized local inversions that successively improve upon a starting model.

As it is conventionally implemented, FWI can be subject to the detrimental effects of cycle skipping. This occurs when some portion of the data predicted by the starting model differs in arrival time by more than half a cycle with respect to the field data. Under these circumstances, FWI will tend to iterate to a local minimum model in which the predicted and observed data are mismatched in time by one or more cycles. Cycle skipping can be difficult to detect, and it generates spurious models; it is the principal reason why FWI requires low-frequency field data and a highly accurate starting model.

Here, we outline a new approach that overcomes the detrimental effects of cycle skipping while retaining the other benefits of conventional FWI. To achieve this, we introduce into FWI some of the beneficial characteristics of wave-equation migration velocity analysis (WEMVA) without introducing its computational burden. In this paper, we outline the methodology and demonstrate its utility using the well-studied 2D Marmousi model. In a companion paper, we apply the new method to a field dataset, and demonstrate that it also performs well for a reflection-dominated dataset.

## Methodology

Conventional FWI seeks to minimise the least-squares difference between observed and predicted datasets. That is, conventional FWI seeks to minimise the least-squares objective function

$$f = \| \mathbf{p}(\mathbf{m}) - \mathbf{d} \|^2 \tag{1}$$

where the column vector **d** represents the field data, and **p** represents the equivalent data predicted by a model **m**. Because seismic data are oscillatory, the objective function given by (1) will typically have many local minima; Figure 1(a) demonstrates this for a field dataset. In this case, local inversion will fail to converge to the global minimum unless the starting model is close to the true model.

The alternative approach that we develop here, employs Wiener filters rather than direct subtraction as the means to compare the two datasets. To do this, we perform the inversion in two stages. In the first stage, we design a convolutional filter **w** that adapts a predicted trace **p** so that it matches an observed trace **d** in a least squares sense, that is $\mathbf{p} * \mathbf{w} \approx \mathbf{d}$. Thus, we find the Wiener filter coefficients **w** that minimise the least-squares objective function $f_1$, where

$$f_1 = \| \mathbf{Pw} - \mathbf{d} \|^2 \tag{2}$$

Here **P** is the convolutional Toeplitz matrix that contains the vector **p** in each column. Equation (2) is applied trace-by-trace. It is linear; it has only one minimum, and provides a unique solution for **w**.

Now, if the filter **w**, determined by equation (2), was simply a delta function at zero lag, then the predicted and observed data would be identical (apart from a scalar). Thus, in a second inversion, we minimise (or maximise) a second normalized least-squares objective function $f_2$, given by

$$f_2 = \frac{\|\mathbf{Tw}\|^2}{\|\mathbf{w}\|^2} \tag{3}$$

where **T** is a simple diagonal matrix that acts to weight the filter coefficients as a function of the magnitude of the temporal lag. If the weights increase with the magnitude of the lag, then $f_2$ should be minimised, and if the weights decrease with lag, then $f_2$ should be maximised.

Minimising (or maximising) equation (3) with respect to the model **m** will push the filter towards a delta function at zero lag, which is equivalent to pushing the predicted data **p** towards the observed data **d**. It is necessary to include the normalisation factor in the denominator in the definition of $f_2$ because otherwise the inversion will tend simply to minimise **w** by increasing the magnitude of **p**.

The inversion scheme defined by (2) and (3) does not suffer from cycle skipping; Figure 1(b) illustrates this for a field dataset. This adaptive waveform inversion scheme does not therefore require an accurate starting model or low frequencies in order to reach the global minimum.

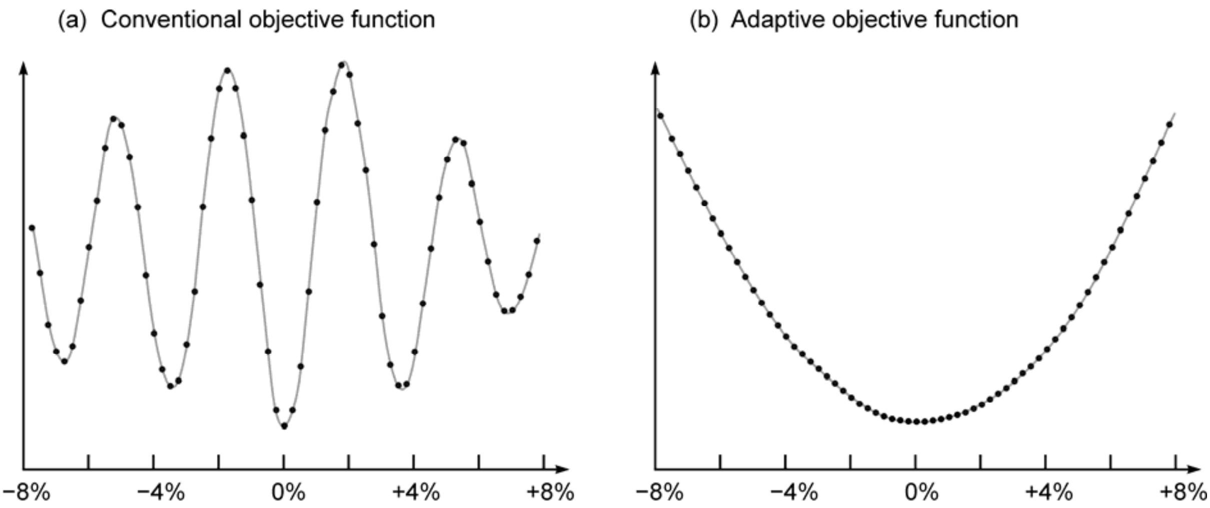


***Figure 1*** *The objective functions given by (a) equation 1, and (b) equation 3, are shown for a single source-receiver pair from a low-pass-filtered 3D field dataset as a function of percentage deviation from a reference velocity model. The dominant frequency of the filtered field data is about 7 Hz; the offset is about 6000 m. The conventional objective function shows cycle skipping when the velocity deviation is more than about 2%; the adaptive objective function is unaffected by such cycle skipping.*

**Relationship to wave-equation migration velocity analysis (WEMVA)**

Adaptive FWI has important features in common with WEMVA; it is these that make it immune to cycle skipping. WEMVA methods typically extend the model in some non-physical way, setting up the inversion so that models evolve towards physical outcomes. In conventional WEMVA, the model is extended by introducing sub-surface offset. This represents non-physical scattering in the subsurface, whereby an incident wavefield at one location generates a coeval scattered wavefield at another location. The inversion is then formulated to focus the energy to zero sub-surface offset, thus producing a physical outcome in which the incident and scattered wavefields are coincident.

Adaptive FWI also extends the model in an analogous non-physical way. The Wiener filters can be regarded as a means of redistributing energy, non-physically, in time. In this case, energy arriving at a receiver at a particular time produces a signal at that receiver at earlier and later times. These non-physical arrivals disappear when the filter becomes a zero-lag delta function; this corresponds to a physical outcome. Conventional WEMVA involves non-physical action at a distance in the sub-surface, whereas adaptive FWI involves non-physical interaction across time at the receivers.

We can now see why the new scheme is immune to cycle skipping. The data **p** predicted by a physical model **m** can be cycle skipped with respect to **d**. However the data predicted by the non-physical extended model (**m** and **w** combined) is the convolution of **w** with **p** which is not cycle skipped since it is always necessarily a close match to **d**. We can therefore regard adaptive FWI as an analogue of WEMVA that seeks to focus energy – all energy and not only primary reflections – at zero temporal lag just as WEMVA seeks to focus primary reflections at zero sub-surface offset.

## Implementation and calculation of the gradient

To perform FWI using this scheme, we must compute the gradient of the objective function $f_2$ with respect to the model parameters $\mathbf{m}$. In the time domain, this leads to a formulation that is computationally similar to conventional FWI. An adjoint source at the receivers is back-propagated for each physical source, and this wavefield is combined with the forward wavefield, and stacked over all sources, to generate the gradient. The adjoint source is obtained by differentiating (3) w.r.t. $\mathbf{p}$.

The method can incorporate all of the techniques commonly employed as part of conventional FWI – composite sources, random subsets of sources, various forms of pre and post-conditioning, and a variety of approximations to the Hessian. The only additional cost and complication is in calculating the adjoint source which involves calculating a 1D Wiener filter for each trace at each iteration, together with some other 1D operations for each trace; the total cost of these is not significant.

Adaptive FWI has two significant advantages over WEMVA. Firstly, it is a scheme that can readily deal with all seismic arrivals including surface and internal multiples, wide-angle refracted and transmitted arrivals; it is not limited only to primary reflections. Secondly, because the extension to the model is imposed only at the receivers, and Wiener filters are computationally efficient, the cost of the method is similar to that of conventional WFI. In contrast, most WEMVA-type schemes are computationally unaffordable for three-dimensional field data because the model extension in the sub-surface typically involves a large additional computational burden in three-dimensional models.

## Application to a synthetic model

To demonstrate the method, we apply it to the well-studied Marmousi model, Figure 2. Data from this model are easy to invert using most FWI schemes provided that the starting model is a smoothed version of the true model and/or that the inversion begins at very low frequencies. In order to provide a realistic test, we here use a one-dimensional starting model that provides only a poor match to the true model, we run the inversion using data that have a dominant frequency of 10 Hz, and we run the inversions throughout using the full data bandwidth without beginning at lower frequencies.

For the demonstration, we use a vanilla version of steepest-descent FWI with no model regularization. We apply trace-by-trace amplitude normalization to both the observed and predicted data, and we spatially precondition the gradient by dividing it locally by the energy in the incident wavefield averaged over all times and all sources. Beyond that, we use no enhancements or additional features.

Figure 2(a) shows the true model and Figure 2(b) shows the starting model. We include free-surface multiples, primary reflections and wide-angle turning arrivals in the data to a maximum offset of about 6 km. At 10 Hz, the majority of the data predicted by the starting model are cycle skipped, and at later travel times, this mismatch is often more than one cycle. Figure 2(c) shows that conventional FWI is unable to recover the correct model at all under these circumstances. Cycle skipping dominates the data, and the recovered model is a poor match to the true model.

In contrast, Figure 2(d) shows the results of applying adaptive FWI to these data using an identical inversion scheme. This method is not affected by the cycle skipping, and it iterates successfully to recover the true model. Its accuracy is affected principally only by the finite bandwidth and finite aperture of the incident wavefield. In the shallowest portion of the recovered model, the result contains some remnant imprint of the sea-bottom, or perhaps of the shortest-offset direct arrival; we anticipate that this minor feature will not be difficult to mitigate.

## Conclusions

Adaptive FWI appears to be capable of recovering the correct global solution in circumstances where conventional FWI fails entirely. The method works by adopting some of the characteristics of WEMVA methods while retaining the essential elements of an FWI scheme, and by doing so it remains computationally efficient. Unlike finite-lag cross-correlation FWI methods, adaptive FWI does not appear to degrade in complex models that involve many interfering events. Unlike some WEMVA-type schemes, adaptive FWI does not require computationally expensive non-zero-lag convolutions in the subsurface, nor does it require the solution of a computationally expensive extended wave equation.

Wiener filters are well understood, and the simple scheme that we have outlined here can, in practice, be extended in many obvious and productive ways to take advantage of that understanding. The filters can for example be varied with time, can be regularized across receivers, and can be multi-dimensional. It is possible to match predicted to observed data, observed to predicted data, and both datasets to each other.

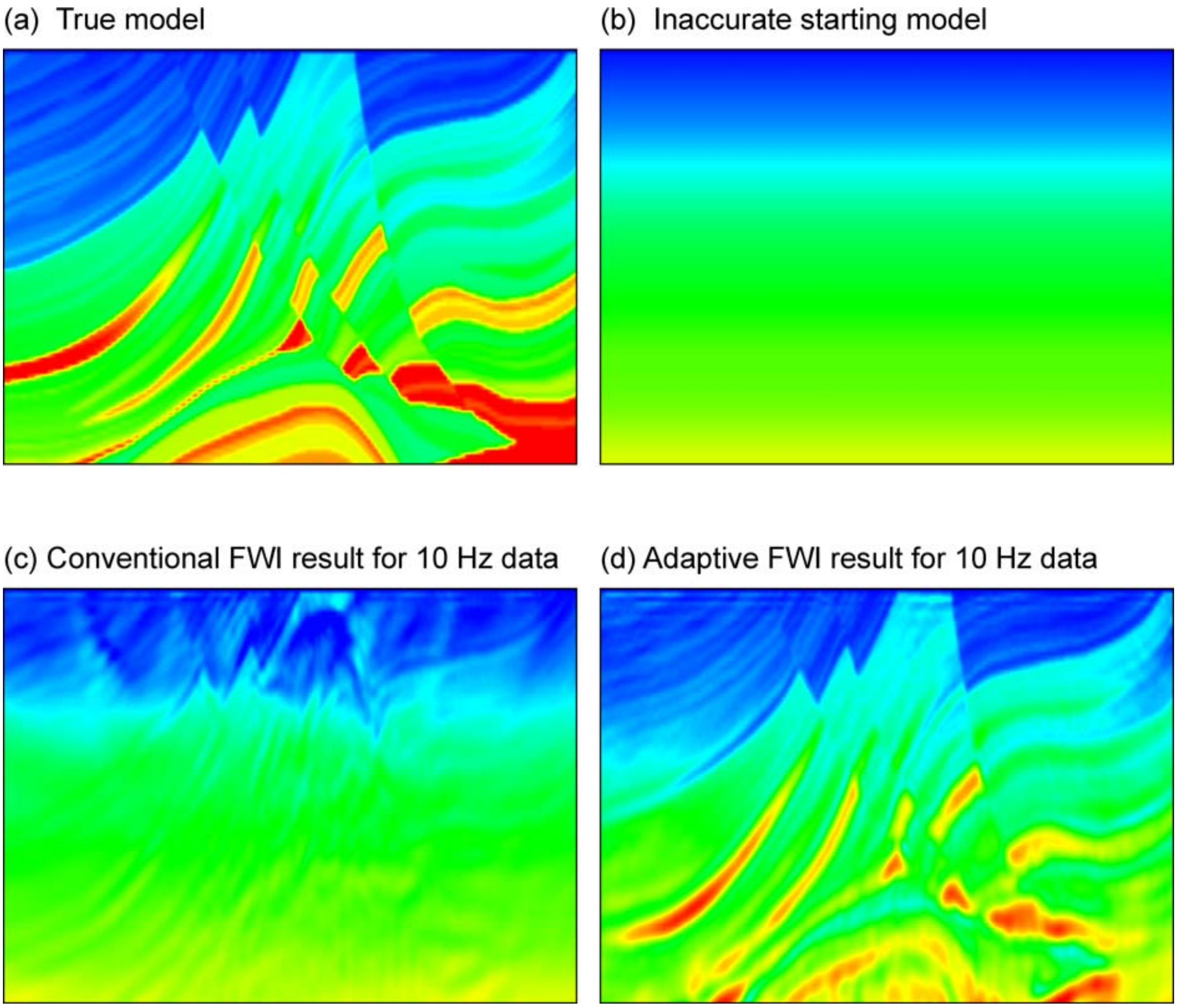


***Figure 2*** *(a) The true velocity model. (b) A one-dimensional inaccurate starting model. (c) Model recovered by conventional FWI. (d) Model recovered using adaptive FWI. At 10 Hz, the data generated by the starting model are badly affected by cycle skipping so that conventional FWI fails entirely while adaptive FWI continues to be accurate and effective.*


**Acknowledgements**

We thank Sub Salt Solutions Ltd for permission to present this paper. The methodology described here is the subject of GB patent application number GB1319095.4.