Th N104 11

Optimised Adaptive Waveform Inversion - Improved Convergence Via Conjugate Gradients and Superior Step-length Calculation

L. Guasch* (Sub Salt Solutions), T. Burgess (Imperial College London) & M. Warner (Imperial College London)

SUMMARY

The insensitivity of Adaptive Waveform Inversion (AWI) to cycle-skipped local minima in the solution space makes it a robust alternative to conventional full waveform inversion when neither good starting models nor low frequencies are available. Despite the absence of such local minima in AWI, steepest-descent optimisation methods used with simple line searches, converge slowly as a result of non-linearity between residuals and model parameters. In this paper, we show that an improved step-length calculation, and use of conjugate gradients, generate shorter paths that require fewer iterations to reach the vicinity of the global minimum, reaching the desired solution much more rapidly, and opening opportunities to apply this technique to raw field data with little prior knowledge of the velocity structure.

Introduction

Conventional full-waveform inversion (FWI) can suffer from cycle skipping. It will tend to become trapped within a local minimum in the objective function if the starting model predicts data that differ by more than half a cycle from the observed data. In 2014, we introduced adaptive waveform inversion (AWI), as a form of FWI that appears to be immune to the effects of cycle skipping (Warner & Guasch, 2014; Guasch & Warner, 2014). Using AWI, it is possible to perform wavefield inversion using an inaccurate starting model that differs significantly from the true model, and such an approach can save the significant time and effort required to build and verify an accurate starting model.

Beginning waveform inversion further from the true model will typically require more iterations to converge. In addition, although the AWI objective function does not contain cycle-skipped local minima, the error surface can be more irregular; especially it appears to contain more long, narrow, trough-like features when the starting model is far removed from the true model, and these act to slow convergence. Presumably the conventional FWI error surface also contains such features, but these are overprinted by the much-stronger effects of cycle skipping. In this paper, we show how the use of conjugate gradients (CG) (Hager & Zhang, 2006), when combined with an improved non-linear step-length calculation, can significantly improve the convergence rate of AWI, and that AWI is then minimally troubled by the sub-optimal topology of the objective function.

Method

Conventional FWI seeks to minimise an objective function formed by the sum of the squares of the sample-by-sample difference in amplitude between the predicted ($\mathbf{d}_{\text{pred}}$) and the observed ($\mathbf{d}_{\text{obs}}$) data:

$$f = \sum_{traces} \frac{1}{2} \|\mathbf{d}_{obs} - \mathbf{d}_{pred}\|^2$$

Since both datasets are oscillatory, this objective function contains local minima when part of one dataset is shifted in time relative to the other by an integer number of cycles. In practice, this means that the starting model must be close to the true model, and the inversion must begin at low frequency, in order for FWI to proceed correctly.

In contrast, AWI uses convolutional matching filters to compare the observed and predicted data. The more closely that these filters resemble zero-lag unit delta-functions, then the more closely that the two datasets will resemble each other. AWI is then configured using an objective function that attempts to focus all the energy in these matching filters to zero lag. If the observed data and the modelling are both noise free, then both AWI and FWI will have the same global minimum, but the FWI objective function will contain many secondary minima that result from cycle skipping whereas AWI will not.

In its simplest formulation, AWI uses one-dimensional linear Wiener filters to perform the matching, it uses separate independent filters for each trace in the dataset, the filter length is comparable to the data length, and the objective function f, is of the form:

$$f = \sum_{traces} \frac{1}{2} \frac{\|\mathbf{T}\mathbf{w}\|^2}{\|\mathbf{w}\|^2}$$

where $\mathbf{w}$ is the Wiener filter that transforms one predicted trace into its equivalent observed trace, and $\mathbf{T}$ is a weighting function that weights each coefficient in $\mathbf{w}$ according to its time shift from zero lag. If $\mathbf{T}$ increases monotonically away from zero lag, then minimising this objective function with respect to the model parameters will act to focus the filters to zero lag, and so act to force the model towards one that predicts data that match the observed data.

Convergence

Although AWI is insensitive to local minima caused by cycle skipping, the shape of the objective function may not be ideal for local optimisation methods so that navigation towards the global minimum can be a slow process that needs many iterations to converge. A simple implementation of the AWI algorithm, using steepest-descent model-updates and a linearly estimated step length, leads to slow convergence rates due to the evaluated functional bouncing around the walls of the many-dimensional valleys in the solution space. There are two reasons why such steepest-descent iterations are inefficient: non-optimal step-length estimation and lack of information from previous iterations in the updates. The first is especially problematic in AWI because the relationship between the model parameters and the residuals, which for AWI consist of normalised weighted Wiener filters, is significantly less linear than in the FWI case, where the Born approximation justifies such a linearisation. As a result of the non-optimal step-length calculation, the ideal scaling of the model update is poorly predicted, and over- and under-estimations of this occur often during the inversion.

A better algorithm for step-length estimation, when combined with conjugate gradients, greatly improves the convergence rate by optimising the search direction at every iteration. This allows the optimisation procedure to make longer and better-directed jumps in the solution space, resulting in a more-efficient evolution that tends to stay at the bottom of its multi-dimensional valleys, and therefore shows a different update direction for each iteration. Figure 1 demonstrates this; it shows velocity updates from a Marmousi inversion with the 1D starting model described in the next section.

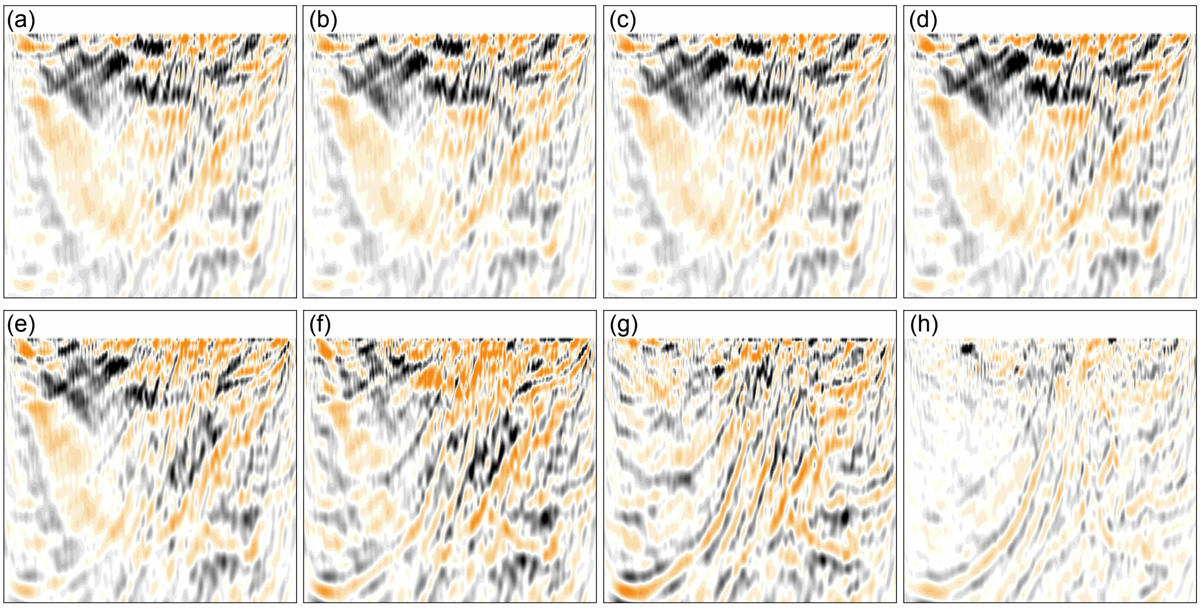

Figure 1 *Normalised velocity updates. Top row: not using conjugate gradients nor the improved line search; bottom row: using conjugate gradients and improved line search. Each column corresponds to the same iteration update: (a) & (e) iteration 2; (b) & (f) iteration 3; (c) & (g) iteration 5; (d) & (h) iteration 10; the first iteration is not shown because it is the same in both cases. The updates evolve significantly more slowly when conjugate gradients and an improved line search are not used, as shown by the similarity of the updates in iterations 2 and 10, figures (a) and (d); in contrast, the updates in the bottom row illustrate how the search direction significantly changes across iterations.*

Synthetic experiment

In order to demonstrate the capabilities of AWI and the benefits of a more sophisticated step-length calculation in combination with CG, we have generated acoustic isotropic synthetic data using the Marmousi model depicted in figure 2(b). We use a starting model that is far removed from the true answer, and this produces cycle skipping for FWI; Figure 2(a) shows the one-dimensional vertical velocity gradient, ranging from 1500 to 3500 m/s, that we used to begin both FWI and AWI. All tests used the following parameters: a 10-Hz peak-frequency source wavelet, 91 sources and 187 receivers in shallow water, spatial preconditioning using an approximate diagonal Hessian, and a total of 500 iterations. The latter is too few without CG, and many more than is required with CG.

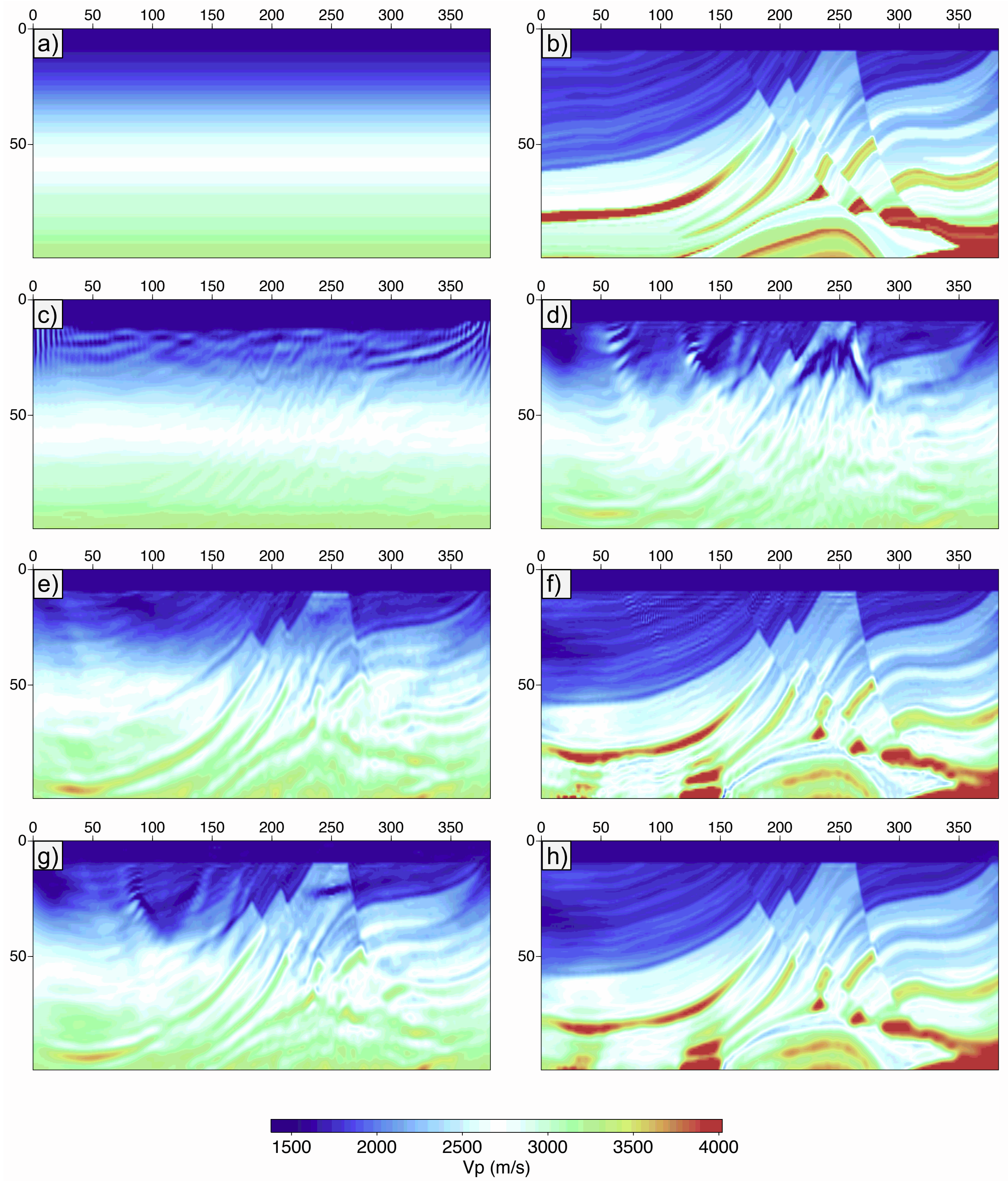

Figure 2 *(a) starting model; (b) true model; (c) FWI inversion without CG; (d) FWI inversion with CG; (e) AWI inversion without CG; (f) AWI inversion with CG; (g) AWI+FWI without CG; (h) AWI+FWI with CG. AWI with CG converges rapidly towards the true solution (f); the shallow artifacts here can be eliminated by smoothing the final model and using it as a starting model for FWI as shown in (h). AWI without CG, on the other hand, is too far from the true model and applying FWI using its final result (e) as a starting model still falls in to a local minimum as in (g).*

FWI clearly fails to converge to the right answer – Figures 2 (c) and 2 (d) – due to the cycle skipping created by the starting model. In this case, the only contribution of the improved line search and CG is to accelerate the convergence towards a local minimum.

Figure 2(e) shows that using AWI, with a linearised step-length calculation and simple steepest descent without CG, is probably heading towards the global minimum and is converging towards the true solution, but after 500 iterations it is still far from the vicinity of this global minimum. This is demonstrated by the fact that this result, if used as the starting model for subsequent FWI inversion, is still significantly cycle-skipped. Figure 2(g) shows a low-velocity hole in the middle of the fault block in the shallow part of cells 230 to 260, and a V-shaped artificial structure between cells 50 and 80; both of these are produced by cycle skipping.

Using CG with an improved line search allows for much faster convergence as shown in Figure 2(f), where the result is close to the true model. When this is used as a starting model for a posterior FWI inversion, it lies within the basin of attraction of the global minimum, and this model is good enough not to generate cycle-skipped data. Note that switching to FWI also eliminates the high-wavenumber shallow artefacts introduced here by the combination of CG and AWI; as yet we do not understand how such artefacts arise, but they are easy to eliminate.

Discussion and Conclusions

As we have previously demonstrated, adaptive waveform inversion succeeds in inverting cycle-skipped data when conventional full-waveform inversion fails entirely. The ability of AWI to circumvent such local minima implies that it is possible to converge from much-more-distant locations in the model space and to ignore any lack of low frequencies in the field data. The higher-frequency data, combined with the non-linearity of the inverse problem, could – and in fact does, at least in the presented synthetic experiment – imply that steepest-descent-based optimisation procedures can require an undesirably high number of iterations to converge to an acceptable solution when starting very far from the true model.

The combination of the CG method with a more accurate step-length calculation – which is important in AWI due to the non-linear relationship between model parameters and residuals – results in more-efficient velocity updates. This combination can greatly improve the convergence rate of AWI, by more than an order of magnitude in some situations. The extra computational cost per iteration of conjugate-gradient methods, and of an improved step-length calculation, is low in comparison to the cost of the large number of extra iterations that are necessary to achieve the same result with a less sophisticated scheme. This approach then makes commercially feasible a workflow in which AWI is used routinely to start from simple velocity models without any requirement for prior detailed velocity-model building. In contrast, if conventional FWI were to be used in this way, it would typically fail catastrophically.

References

Guasch, L., and Warner, M., 2014, Adaptive Waveform Inversion – FWI without cycle skipping: Practice, 76th EAGE Conference & Exhibition, Extended Abstracts, Amsterdam.

Hager, W., and H. Zhang, 2006, A survey of nonlinear conjugate gradient methods. Pacific journal of Optimization 2.1: 35-58.

Warner, M., and L. Guasch, 2014, Adaptive Waveform Inversion – FWI without cycle skipping: Theory, 76th EAGE Conference & Exhibition, Extended Abstracts, Amsterdam.