Adaptive Traveltime Inversion: Mitigating cycle-skipping by minimization of the moment of the matching filter distribution

March 5, 2020

1 Summary

We present a method to obtain a misfit function for robust waveform inversion. This method, designated as adaptive traveltime inversion (ATI), computes a matching filter that matches the measured data to the predicted one. If the velocity model is accurate, the resulting matching filter reduces to an (approximate) Dirac delta function. Its traveltime shift, which characterizes the defocusing of the matching filter, is computed by minimization of the cross-correlation between a penalty function like $t^2$ and the matching filter. The ATI misfit function is constructed by the minimization of the least square error of the calculated traveltime shifts. Further analysis shows that the resulting traveltime shifts correspond to the first-order moment, the mean value, of the resulting matching filter distribution. We extend ATI to a more general misfit function formula by computing different order moment of the resulting matching filter distribution. Choosing the penalty function in adaptive waveform inversion (AWI) as $t^2$, the misfit function of AWI can be interpreted as the second-order moment. We use the Marmousi model to verify the effectiveness of the proposed method.
2 Introduction

Designing a robust misfit function, which can mitigate cycle skipping is crucial for Full-Waveform Inversion (FWI). The conventional $l_2$ norm misfit function though simple, and has the potential in delivering high resolution models, is a local comparison between the data and is prone to cycle skipping. Thus, many have addressed the cycle-skipping problem by developing more robust misfit functions, which try to compare the data more globally such as the matching filter based misfit function $[?, ?, ?, ?]$.

The matching filter approach has been in the literature for a while. It dates back to the early work of wave-equation traveltime inversion (WTI) by $[?]$. Their approach uses the cross-correlation to compute the traveltime shift between the modeled and measured seismic data. The misfit function was formulated as the least square errors of those traveltime shifts. Compared with conventional ray-based traveltime tomography, their method is an automatic wave-equation based inversion method free of traveltime picking. However, in their approach, the cross-correlation is applied to the seismic data directly, with an inherent limitation of the cross-correlation in which it focuses on dominant events with high amplitude. As a result, their approach tends to be low resolution compared to conventional FWI $[?]$. Improvements have been made by computing a matching filter between the predicted data and the measured one, based on deconvolution instead of cross-correlation. If the velocity model is accurate, the non-zero coefficients of the filter will focus at the zero lag. A misfit function can be designed by penalizing the coefficients of the filter at the none-zero lag to update the model $[?]$. $[?]$ introduced adaptive waveform inversion (AWI) by adding an extra normalization term and in their study, the normalization term plays an important role of providing better convexity for the misfit function and accelerating its convergence. This normalization term has its roots in the instantaneous traveltime objective function as demonstrated by $[?]$.

In this abstract, we propose a method named adaptive traveltime inversion (ATI), to overcome the limitation of the cross-correlation operator used in the WTI method. As mentioned above, applying cross-correlation between the measured and predicted data directly would focus the inversion on the dominant events with high amplitude. This makes the WTI method admit low-resolution results and limits its application. To resolve this issue, we propose applying the cross-correlation between the penalty function $t^2$ and the resulting matching filter computed from the measured and predicted data. The resulting traveltime shifts, which minimize the cross-correlation, would indicate the defocusing of the matching filter. The least squares error of the resulting traveltime shift can be formulated as a misfit function. We will generalize our result and show that the traveltime shifts in ATI have a clear physical meaning and it is the first-order moment, i.e., the mean value, of the corresponding matching filter distribution while the misfit function of AWI is the second order moment. This kind of generalization allows us to better understand the relation between ATI and AWI misfit functions. It also provides insights into designing new misfit functions such as using high-order moments of the matching filter distribution.
We present the results for the Marmousi model to demonstrate the good performance of the proposed method.

3 Wave-equation Traveltime Inversion

The method presented here is partially inspired by the wave-equation traveltime inversion (WTI). So we start with a brief review of the method [7]. WTI tries to minimize the misfit function defined by the square errors of the traveltime shift $\Delta \tau$:

$$J_{\text{WTI}} = \frac{1}{2} \Delta \tau^2.$$  \hfill (1)

The traveltime shift $\Delta \tau$ corresponds to the time shift, which attains the maximum value for the cross-correlation function $f_{\text{WTI}}(\tau)$ between predicted and measured data, i.e.,

$$f_{\text{WTI}}(\tau) = \int (t + \tau)p(t)dt,$$  \hfill (2)

where $d(t)$ is the measured data and $p(t)$ is the predicted data computed using the current velocity model.

4 Adaptive Traveltime Inversion

For adaptive traveltime inversion (ATI), our misfit function uses the exact form as WTI of equation (1):

$$J_{\text{ATI}} = \frac{1}{2} \Delta \tau^2.$$  \hfill (3)

However, the time shift $\Delta \tau$ is computed in a different way. We would at first compute a matching filter $w(t)$, which matches the observed data to the predicted data : $d(t) * w(t) = p(t)$, where $*$ denotes the convolution operation. If the velocity model for producing the predicted data $p(t)$ is correct. The resulting matching filter should end up being an (approximate) Dirac delta function. Otherwise, its coefficients will spread out over non-zero lags. Similarly, we define a cross-correlation function $f_{\text{ATI}}(\tau)$ to measure such defocusing:

$$f_{\text{ATI}}(\tau) = \frac{1}{2} \int (t + \tau)^2w^2(t)dt,$$  \hfill (4)

the time shift $\Delta \tau$ defined in the ATI misfit function of equation (3) corresponds to the minimum value of $f_{\text{ATI}}(\tau)$. It is obvious that if the velocity is correct, the matching filter $w(t)$ would be an (approximate) Dirac delta function, with the minimum value corresponding to time shift $\Delta \tau = 0$. Otherwise, the time shift $\Delta \tau$ is nonzero and by minimizing the square errors of equation (3), we can update the velocity model.

As $\Delta \tau$ corresponds to the minimum value of $f_{\text{ATI}}(\tau)$, its first-order derivative at the minimum corresponding traveltime shift would be zero, thus, we use the
corresponding connective function, given by:

\[
\frac{df_{\text{ATI}}(\tau)}{d\tau}|_{\tau=\Delta\tau} = \int (t + \Delta t)w^2(t)dt = 0. \tag{5}
\]

By a simple arithmetic operation, we can express the time shift \(\Delta\tau\) in equation (5) as:

\[
\Delta\tau = -\int tw^2(t)dt = -\frac{\int tw^2(t)dt}{||w||_2^2}. \tag{6}
\]

Substituting the time shift \(\Delta\tau\) in equation (5) into equation (2), we have the final explicit formula for the ATI misfit function:

\[
J_{\text{ATI}} = \frac{1}{2} \left[ \frac{\int tw^2(t)dt}{||w||_2^2} \right]^2. \tag{7}
\]

5 Discussion

5.1 ATI and WTI

Comparing ATI with WTI, the main difference is the cross-correlation functions of equations (2) and (2). In WTI, the cross-correlation function of equation (2) correlates the observed and predicted data directly, and this kind of direct cross-correlation between waveforms makes it focus on the events with large amplitudes, and thus, prevents WTI from handling properly scattered energy in the data as such scattered energy normally has small amplitudes. As reported by [?], WTI has relatively low resolution compared to full waveform inversion [?].

For ATI, the cross-correlation strategy is a bit different: a matching filter \(w(t)\) is computed at first. The matching filter provides a global comparison between the predicted and observed data, and this feature makes the resulting ATI method handles the cycle-skipping issue, and thus, make it more "adaptive." The ATI cross-correlation of equation (2) plays the role of focusing the resulting matching filter and forces it to be a Dirac Delta function. Unlike the cross-correlation function of equation (2) for WTI, equation (2) of ATI does not cross-correlate the predicted and observed data, which has the potential drawback of focusing on events with high amplitude. ATI uses cross-correlation to measure the defocusing of the matching filter, and as it is a global approach, it can handle the full wavefield in the data including direct arrivals, transmissions, refractions, reflections, etc. Although "traveltime" is included in the name of "ATI", we will show in the following examples, it is actually a high-resolution full-waveform inversion method. We include "traveltime" in the name to highlight its potential similarity with the wave-equation traveltime inversion method.
5.2 ATI and AWI

In adaptive waveform inversion (AWI)\cite{[citation]}, if we choose the penalty function as $T(t) = |t|$, the misfit function of AWI is given by:

$$J_{AWI} = \frac{1}{2} \frac{\int t^2 w^2(t) dt}{||w||_2^2}$$ \hspace{0.5cm} (8)

if we compare equation (8) with equation (7), we can see that they are generally similar. We can express them under a more general form based on the n-th order moment of a probability distribution,

$$\mu_n = \int t^n q(t) dt,$$ \hspace{0.5cm} (9)

where $q(t)$ is the probability distribution computed from matching filter: $q(t) = \frac{w^2(t)}{\int w^2(t) dt}$.

Thus, both ATI and AWI try to measure the focusing of the resulting matching filter by computing its moment. For ATI, it is the first-order moment, and for AWI, the second-order moment. A Higher-order moment with $n > 2$ may also be selected as a misfit function. For example, the third-order moment, the skewness $\mu_3$, is a measure of the lopsidedness of the distribution, while the fourth order moment $\mu_4$, the kurtosis, is a measure of the heaviness of the tail of the distribution. As the moment of a distribution is a statistical measure. Other statistical measures, such as geometrical mean or median, may also be used as a misfit function as well, and those optional misfit functions may highlight additional features.

6 Results

As shown in Figure 1a the true velocity $v_{true}$ of the Marmousi model extends 2 km in depth and 8 km laterally. The initial velocity $v_{ini} = v(z)$ is shown in Figure 1b. The source wavelet is a Ricker wavelet with a 10 Hz peak frequency. In the inversion, we mute data below 3 Hz to verify that our proposed method is capable of overcoming the cycle skipping problem without low frequency. In these time domain inversions, we apply a one-time low pass filter to the data up to 10 Hz. Therefore, we do not sequentially increase the frequency as often done. Instead, we perform a full band inversion for the frequency range of 3 Hz to 10 Hz and no window muting or any damping are applied to the data. This is a challenge for the conventional least squares $l_2$ misfit function as for this frequency band we face the cycle-skipping issue. The inverted result for the conventional least squares $l_2$ norm misfit function is shown in Figure 1c. Due to cycle skipping, the result has very strong artifacts in the left part of the model, and it is generally far from the true model. Figure 1d is the result of using the WTI misfit function. We can see that it can recover the shallow part of the velocity model well. However the deeper reflectors are not well positioned, and
the model is overall low resolution. Figures 1e and 1f show results from ATI and AWI methods. We can see that both methods recover the model well, both the low wavenumber and fine-scale parts of the true model although there are some minor artifacts at the boundary due to limited illumination. Considering the initial velocity is far away from the true one, the proposed ATI method like AWI can reduce the cycle-skipping issue and converge towards the global minimum successfully.

7 Conclusions

We proposed a robust misfit function for FWI which we refer to as adaptive traveltime inversion. The inversion is formulated by minimizing the least square errors of the traveltime shifts computed from the cross-correlation of a penalty function with a matching filter, which is computed by a deconvolution of the measured and computed data. Compared to conventional least square $l_2$ norm misfit function where we perform a sample-by-sample comparison between the observed and modeled data, ATI admits a global trace by trace comparison. Since the traveltime shift is not computed using cross-correlation of the measured and predicted data directly, ATI admits higher accuracy and resolution than the WTI method. The proposed method also allows us to extend the misfit function to a more generalized form in terms of moments of the resulting matching filter distribution: Specifically, the proposed misfit function can be considered as a (squared) first-order moment of the matching filter distribution. While AWI can be considered as a special case of the second-order moment, higher-order moments such as skewness (third order) or kurtosis (fourth order) can also be used as misfit functions.

8 Acknowledgements

We thank KAUST for supporting this research and members of SWAG for useful discussions.

References


Figure 1: a) The true velocity; b) The initial velocity: inverted velocity by c) least squares $l_2$ norm; d) WTI; e) ATI; f) AWI for the modified Marmousi model.