Adaptive Traveltime Inversion with information entropy regularization

Bingbing Sun and Tariq Alkhalifah, King Abdullah University of Science and Technology

SUMMARY

A recently introduced Adaptive Traveltime Inversion (ATI) provided us with a robust misfit function for reducing cycle skipping in Full-Waveform Inversion (FWI). Unlike the conventional L_2 -norm approach, ATI computes a matching filter first by deconvolution of the predicted data with the measured ones. If the velocity model is relatively accurate, the resulting matching filter is close to a 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. ATI is constructed by minimization of the least square errors of the calculated traveltime shifts. It has been shown that the resulting traveltime shift corresponds to a first-order moment, which corresponds to the mean value of the resulting matching filter distribution. In order to accelerate the convergence and improve the robustness of the ATI approach, in this abstract, we propose to constraint the variance of the resulting matching filter using an information entropy function. We further demonstrate that, in comparison to AWI which tries to minimize the sum of the mean and the variance of the resulting matching filter, the misfit of ATI with entropy minimizes the sum of the mean and the logarithm of the variance instead. We use the Marmousi example and an offshore field dataset to demonstrate the effectiveness of the proposed method.

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 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 (Luo and Sava, 2011; Warner and Guasch, 2016; Sun and Alkhalifah, 2018a, 2019b,c).

Matching filters have been used in seismic processing and analysis in many applications ranging from noise removal to source estimation (reference). In waveform inversion, a simplification of it computed based on crosscorrelation was adopted by Van Leeuwen and Mulder (2008). The full use of a matching filter in waveform inversion can be attributed to Luo and Sava (2011), in which their optimization problem minimizes energy away from the zero-lag region of the matching filter. However, their optimization misses a crucial normalization term adopted by Warner and Guasch (2016) that helps focusing the energy to zero lag.

Sun and Alkhalifah (2019a) proposed a method named Adaptive Traveltime Inversion (ATI), that measures the phase dis-

tance of the matching filter from zero (A Rytov implementation). Applying cross-correlation between the measured and the predicted data directly would focus the inversion on the dominant events with high amplitude. This makes the such methods admit low-resolution results and limits its application. A remedy can be made by 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. Sun and Alkhalifah (2019a) generalize the 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 adaptive waveform inversion (Warner and Guasch, 2016) is the second order moment.

In this abstract, we proposed to accelerate the convergence and the robustness of the ATI method with an extra regularization term which applies an information entropy function to the resulting matching filter. The introduced entropy regularization term will constrain the variance of the resulting matching filter. Together with the original ATI misfit, the new misfit function with entropy regularization will physically force the resulting matching filter to evolve to zero-time lag and at the same time become more compact. Mathematically, it achieves this by minimization of the sum of the mean and logarithm of the variance of the resulting matching filter as a probability distribution. In this sense, it has similarity with adaptive waveform inversion (AWI), which tries to minimize the sum of the mean and the variance of the matching filter instead. We use the Marmousi example and a field dataset from Australia to test the proposed method.

ADAPTIVE TRAVELTIME INVERSION

Here, we briefly review the ATI approach (Sun and Alkhalifah, 2019a). The ATI misfit function admits the form of

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

Unlike in wave equation traveltime inversion (Luo and Schuster, 1991) where the traveltime shift is computed by cross-correlation of the predicted data and the measure data directly, here the time shift $\Delta\tau$ is computed in a different way. We would 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 func-

tion $f_{\text{ATI}}(\tau)$ to measure such defocusing:

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

the time shift $\Delta\tau$ defined in the ATI misfit function of equation 1 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 1, 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:

$$\left. \frac{df_{\text{ATI}}(\tau)}{d\tau} \right|_{\tau=\Delta\tau} = \int (t + \Delta\tau) w^2(t) dt = 0. \quad (3)$$

Using a simple arithmetic operation, we can express the time shift $\Delta\tau$ in equation 3 as :

$$\Delta\tau = - \frac{\int t w^2(t) dt}{\int w(t)^2 dt} = - \frac{\int t w^2(t) dt}{\|w\|_2^2}. \quad (4)$$

Substituting the time shift $\Delta\tau$ in equation 4 into equation 1, we have the final explicit formula for the ATI misfit function:

$$J_{\text{ATI}} = \frac{1}{2} \left[\frac{\int t w^2(t) dt}{\|w\|_2^2} \right]^2. \quad (5)$$

INFORMATION ENTROPY REGULARIZATION

We introduce the matching-filter probability distribution $\tilde{w}(t) = w^2(t)/\|w\|_2^2$, which fulfills the requirement of a probability distribution, i.e., it is nonnegative and summation equals to one. We can recognize that the ATI misfit function of equation 5 is actually the mean value of the matching filter distribution. In order to make the resulting matching filter compact, we introduce the following information entropy regularization term (Carter, 2007):

$$R_{\text{entropy}}(\tilde{w}(t)) = - \int \tilde{w}(t) \ln(\tilde{w}(t)) dt. \quad (6)$$

As shown in the Appendix, the entropy regularization term actually tries to minimize the logarithm of the variance of the resulting matching filter. Thus, the final misfit function can be formulated as

$$J_{\text{new}} = \left[\int t \tilde{w}(t) dt \right]^2 - \lambda \int \tilde{w}(t) \ln(\tilde{w}(t)) dt, \quad (7)$$

where λ is a weighting term which balances the ATI misfit and the entropy regularization term.

Considering the matching filter distribution $\tilde{w}(t)$ is approximated by a Gaussian distribution with mean μ and variance σ^2 . The misfit function of equation 7 can be evaluated as (see the appendix for the derivation):

$$J_{\text{new}}(\mu, \sigma^2) = \mu^2 + \frac{\lambda}{2} [\ln(2\pi\sigma^2) + 1]. \quad (8)$$

As shown by Sun and Alkhalifah (2019c), the application of the AWI misfit (Warner and Guasch, 2016) to a Gaussian distribution leads to:

$$J_{\text{AWI}}(\mu, \sigma^2) = \mu^2 + \sigma^2. \quad (9)$$

Comparing of Equations 8 and 9, we can see that our and the AWI misfits show similar objectives of focusing the resulting matching filter to an approximated Dirac delta function by minimization of its mean and variance. The difference is that our misfit function utilizes the information entropy function to control the compactness of the resulting matching filter distribution while AWI does it by minimizing the variance directly. More importantly, the new objective mildly penalizes the large variances compared with AWI, which will help in focusing the near zero lag energy first, often corresponding to the shallower events. Usually, there are bigger mismatches between the measured and predicted data at later arrivals.

EXAMPLES

In this section, we apply our approach to invert for the modified Marmousi model. The true velocity v_{true} shown in Figure 1a extends 2 km in depth and 8 km, laterally. The initial velocity is shown in Figure 1b. The dataset is modeled using 80 shots with a source interval 100 m and 400 receivers with an interval of 20 m. The source wavelet is a Ricker wavelet with a 10 Hz peak frequency. We mute data below 3 Hz to verify that our proposed method is capable of overcoming the cycle skipping for data free of low frequencies. In the inversion, we do not perform frequency continuation. Instead, we band-pass the dataset with a high cutting frequency of 10 Hz and perform the inversion directly. We run the inversion for 100 iterations, and the final inverted result for the L2 norm, ATI, ATI with entropy and AWI misfits are shown in Figures 1c to 1f, respectively. From the results, the L2-norm misfit fails to capture the true Marmousi in many areas due to cycle skipping especially at depth while ATI with entropy regularization results in a mildly better model than AWI (look at the red arrows) and can mitigate the cycle-skipping and outperform the result from the ATI misfit with better focusing (see the area pointed by black arrows).

The second example is a marine real data set from offshore Australia (Sun and Alkhalifah, 2018b). The offset range is from 160 m to 8200 m. The initial vertical velocity is converted from RMS velocity given in Figure 2a, and we set the anisotropic parameter $\eta = \delta = 0$ for the initial model. We perform the VTI-FWI inversion using the proposed ATI misfit with entropy regularization with a low-pass filter applied to the data from 4 Hz to 9 Hz every 1 Hz sequentially and refine the obtained result using the L2-norm misfit function at the end for higher resolution up to 24 Hz. The final inverted model for the nmo velocity, η and δ is shown in Figures 2b to 2d, respectively. We can see that the inverted velocity model shows consistent structures. In the right panels of Figures 3a and 3b, we show one selected common shot gather for the initial and the inverted models. We compare them with the recorded shot gather at the same location in the right panel. Clearly, the inverted model reproduces the data that better matches the mea-

sured data, especially at the larger offsets where cycle skipping usually happens, and it is evident for the initial model. Considering the initial velocity model is obtained from a crude RMS velocity analysis, we attribute the good convergence of the proposed misfit function to its ability to handle cycle skipped data. In Figures 3c and 3d, we compare the RTM image for the initial model and the inverted model. It is clear that using the VTI-FWI inverted model, the image shown in Figure 3d is better focused (note the areas denoted by the blue arrows).

CONCLUSION

We introduce an entropy regularization term to the ATI misfit to accelerate the convergence and improve the robustness. The extra entropy regularization term will try to focus the resulting matching filter by minimization its variance. The Marmousi example and a field dataset from Austria demonstrated the effectiveness of the proposed method.

ACKNOWLEDGMENTS

We thank the members of SWAG group for useful discussions.

APPENDIX A

ENTROPY OF A GAUSSIAN DISTRIBUTION

Below, we show that the entropy given by Equation 6 computes the logarithm of the variance of a Gaussian distribution.

The Gaussian probability density function is defined by: $N(x; \mu, \sigma) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-(x-\mu)^2/2\sigma^2}$. All probability density function must be normalized to unity, thus, we have:

$$\int N(x; \mu, \sigma) dx = 1. \quad (\text{A-1})$$

The expectation of $(x - \mu)^2$, $E(x - \mu)^2$, i.e, the variance can be evaluated as:

$$E(x - \mu)^2 = \int (x - \mu)^2 N(x; \mu, \sigma) dx = \sigma^2. \quad (\text{A-2})$$

By Equation 6, the entropy of the Gaussian distribution can be computed as:

$$\begin{aligned} \text{Entropy}(\mu, \sigma^2) &= \\ &= - \int \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-(x-\mu)^2/2\sigma^2} \ln \left[\frac{1}{(2\pi\sigma^2)^{1/2}} e^{-(x-\mu)^2/2\sigma^2} \right] dx \\ &= \frac{1}{2} \ln(2\pi\sigma^2) \int (2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2} dx \\ &+ \frac{1}{2\sigma^2} \int (2\pi\sigma^2)^{-1/2} (x - \mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx. \end{aligned} \quad (\text{A-3})$$

Introducing equation A-1 and equation A-2, it can be modified as :

$$\text{Entropy}(\mu, \sigma^2) = \frac{1}{2} \ln(2\pi\sigma^2) \cdot 1 + \frac{1}{2\sigma^2} \cdot \sigma^2 = \frac{1}{2} [\ln(2\pi\sigma^2) + 1]. \quad (\text{A-4})$$

Figure 1: a) The true velocity; b) the initial velocity of the modified Marmousi model; The inverted velocity for c) L2-norm misfit; d) ATI misfit; e) ATI with entropy misfit; f) AWI misfit.

Figure 2: The initial model for a) the nmo velocity; the inverted model for b) the nmo velocity; and the anisotropic parameters c) η and d) δ .

Figure 3: The record comparison of the measured data (left panel) and the modeled data by a) the initial model (right panel); b) the inverted model (right panel); the RTM image for c) the initial model; d) the inverted model.

REFERENCES

- Carter, T., 2007, An introduction to information theory and entropy: Complex Systems Summer School, SantaFe, 22–34.
- Luo, S., and P. Sava, 2011, A deconvolution based objective function for wave equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792.
- Luo, Y., and G. T. Schuster, 1991, Wave equation travelttime inversion: *Geophysics*, **56**, 645–653.
- Sun, B., and T. Alkhalifah, 2018a, The application of an optimal transport to a preconditioned data matching function for robust waveform inversion: SEG Technical Program Expanded Abstracts, 5168–5172.
- , 2018b, Automatic wave-equation migration velocity analysis by focusing subsurface virtual sources: *Geophysics*, **83**, no. 2, U1–U8.
- , 2019a, Adaptive travelttime inversion: mitigating cycle-skipping by minimization of the moment of the matching filter distribution: 81st EAGE Conference and Exhibition.
- , 2019b, A robust waveform inversion using a global comparison of modeled and observed data: *The Leading Edge*, **38**, no. 3, 185–192.
- , 2019c, Salt body inversion using an optimal transport of the preconditioned matching filter: 81st EAGE Conference and Exhibition.
- Van Leeuwen, T., and W. Mulder, 2008, Velocity analysis based on data correlation: *Geophysical Prospecting*, **56**, no. 6, 791–803.
- Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: *Geophysics*, **81**, no. 6, R429–R445.