GEOPHYSICS®

Adaptive Traveltime Inversion

Journal:	<i>Geophysics</i>
Manuscript ID	GEO-2018-0595.R2
Manuscript Type:	Technical Paper
Keywords:	full-waveform inversion, crosscorrelation, deconvolution
Area of Expertise:	Seismic Velocities and Traveltimes, Seismic Inversion

SCHOLARONE™
Manuscripts

Adaptive Traveltime Inversion

Bingbing Sun and Tariq Alkhalifah

bingbing.sun@kaust.edu.sa

tariq.alkhalifah@kaust.edu.sa

Physical Sciences and Engineering

King Abdullah University of Science and Technology

Thuwal, Saudi Arabia

(April 20, 2019)

Running head: **Adaptive Traveltime inversion**

ABSTRACT

We present a method to obtain a misfit function for robust waveform inversion. In this method, designated as adaptive traveltime inversion (ATI), a matching filter that matches predicted data to measured data is computed. 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 shift. Further analysis shows that the resulting traveltime shift corresponds to a 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 is the second order moment, the variance of the resulting matching filter distribution with zero mean. Since the proposed ATI method is based on a global comparison using deconvolution, like AWI, it can resolve the "cycle skipping" issue. We illustrate the proposed ATI misfit function and compare it with state of the art options such as least squares inversion (L2 norm), wave-equation travelttime inversion (WTI), and AWI using schematic examples before moving to more complex examples, like the Marmousi model. For the Marmousi model, starting with a 1D $v(z)$ model, with data without low frequencies (no energy below 3 Hz), a meaningful estimation of the P-wave velocity model is recovered. The proposed ATI misfit function (first-order moment) shows comparable performance to the AWI misfit function (the second order moment). We also include a real data example from the Gulf of Mexico to demonstrate the effectiveness of the proposed method.

INTRODUCTION

Full Waveform Inversion (FWI) (Virieux and Operto, 2009) is a high-resolution velocity model building methodology. It is a nonlinear inversion process and thus, we iteratively and sometimes slowly update the subsurface model to reduce the mismatch between the predicted and measured seismic data. In conventional seismic data processing and imaging, we normally use parts of the information embedded in the seismic data, such as first arrival traveltimes in near-surface tomography. FWI aims at using the whole recorded waveform signal, and all the seismic events (diving waves, pre-and post-critical reflections, and converted waves) are taken into account to invert for the physical properties that produced and characterized these events, such as P- and S-wave velocities, attenuation, density and anisotropy parameters. FWI normally yields high-resolution estimation of the medium compared to tomographic methods, up to the theoretical limit of half of the shortest wavelength of the recorded signal. Although the basic concept and theory of FWI in the time domain was introduced and developed by Lailly (1983) and Tarantola (1984), there are variations of the FWI formulation such as frequency domain FWI (Pratt, 1999; Sirgue and Pratt, 2004), Laplace domain and Laplace-Fourier domain FWI (Shin and Cha, 2008, 2009). Due to recent advances of super computers, long-offset wide-azimuth, and broadband seismic sources, FWI gained renewed interest in oil and gas exploration and we managed to apply it to real data in a more consistent fashion. There are numerous reported successful applications to 3D and real data (Operto et al., 2006; Plessix and Perkins, 2010; Sirgue et al., 2010; Warner et al., 2013; Vigh et al., 2014; Shen et al., 2018; Alkhalifah et al., 2018; Zhang et al., 2018).

Because of the nature of this nonlinear inverse problem, "cycle skipping" is a major issue.

If the initial model is far from the true one or some portion of the predicted data differs in phase by more than a half cycle of the dominant frequency in the measured data, the inversion often gets stuck in a local minimum. One solution for reducing "cycle skipping" is feeding FWI with relatively accurate initial model. Currently, the standard velocity building tools in the industry highly rely on ray-based tomographic methods (Woodward et al., 2008). The resulting velocity model using the ray-based approach tends to be low resolution and less accurate. In complex media, ray-based tomography often fails due to multi-pathing and the high-frequency asymptotic assumption utilized in the approach (Wang et al., 2008). Performing FWI sequentially with increasing offset or frequency can be another way of resolving the cycle skipping problem (Sansigre and Ratcliffe, 2014). However, the application of offset and frequency control in FWI needs skillful QC and intensive manual interactions of the inversion process and increases the turnaround time significantly especially for the 3D project. An ideal solution to the "cycle skipping" would be designing a more convex misfit function, which can adaptively control the inversion process and allows the model to gradually converge to the exact one even starting from a far away, "cycle-skipped" model. This kind of more robust misfit function can thus, allow us to perform such as full-band and full-offset waveform inversion with limited manual interactions (Debens et al., 2017).

It is well known that the least squares (L2-norm) misfit function is prone to "cycle skipping" due to its inherent limitation of local, sample by sample comparison. One potential solution to the "cycle skipping" issue in FWI is to use global optimization schemes such as the Monte Carlo method (Jin and Madariaga, 1994; Sambridge and Mosegaard, 2002), genetic algorithms (Sen and Stoffa, 1992; Jin and Madariaga, 1993), simulating annealing (Kirkpatrick et al., 1983; Datta and Sen, 2016). However, global optimization schemes

typically require an evaluation of the misfit function for tens of thousands of times. For real applications, especially in 3D, where we have hundreds of millions of unknown parameters, these strategies become impractical considering the current computational capabilities. Thus, current FWI methods rely on local optimization schemes in which we mainly update the model along the descent directions.

In addition to the data-domain misfit functions, image-domain misfit functions based on the concept of "Differential Semblance Optimization" (Symes and Carazzone, 1991) have provided potential solutions for the cycle skipping dilemma. In these methods, we assume that the velocity can be separated into low wavenumber background and high wavenumber reflectivity. Based on the extended imaging condition, an extended image (reflectivity) can be computed and used to evaluate the accuracy of the background velocity model. The misfit function can be defined to measure the focusing of the energy in the subsurface offsets (Shen and Symes, 2008; Sun and Alkhalifah, 2018b) and the time lag gathers (Sava and Fomel, 2006) or measuring the coherency of the events in the angle domain gathers (Biondi and Symes, 2004). Although this category of methods show more robustness in dealing with "cycle skipping," the computation of the extended images can increase the computational burden significantly. In the meantime, as these methods assume that only primary reflections are used to generate the images, considerable preprocessing is needed to meet this requirement.

Thus, many researchers have addressed the cycle-skipping problem by introducing more robust misfit functions, which try to compare the data in a more global way such as matching filter based misfit function (Van Leeuwen and Mulder, 2008, 2010; Luo and Sava, 2011; Warner and Guasch, 2016) and optimal transport misfit function (Engquist and Froese, 2014; Engquist et al., 2016; Métivier et al., 2016; Yang et al., 2018; Yang and Engquist,

2018; Qiu et al., 2017; Sun and Alkhalifah, 2019). As the proposed method in this paper is related to the matching filter approach. Thus, we will give a review of the category of misfit function designing using matching filter.

The matching filter based approach has been in the literature for a while. It can be dated back to the early work of wave-equation traveltime inversion (WTI) by Luo and Schuster (1991). In their approach, they use the cross-correlation to compute the traveltime shift between the predicted and measured seismic data. They formulate a misfit function based on the least square errors of those traveltime shifts. Compared with conventional ray-based traveltime tomography, their method is an automatic wave-equation based method free of traveltime picking. However, in their approach, the cross-correlation is applied to the seismic data directly, and the inherent limitation of cross-correlation indicates that it will focus on dominant events with high amplitude. As a result, their approach tends to be low resolution compared to conventional FWI (Zhou et al., 1995, 1997; Zheng et al., 2013).

The idea behind the matching filter approach is that instead of comparing the seismic trace sample by sample such as subtraction in the least squares L2 norm, we would at first compute a matching filter by deconvolving the measured data from the predicted data (or vice versa). If the velocity used in modeling is relatively accurate, the resulting matching filter would have its energy focused at zero time-lag and become an (approximated) band limited Dirac delta function. Otherwise, by penalizing the coefficients at non zero time-lag, we can formulate an optimization problem. Luo and Sava (2011) initially proposed this idea of designing a misfit function for FWI using matching filters computed per trace. In their approach, they designed a misfit function based on deconvolution instead of cross-correlation (Van Leeuwen and Mulder, 2008, 2010). Thus, the objective function becomes less sensitive to band limited or non-impulsive source functions, which may result in a

nonzero gradient of the objective function even when the inverted velocity model matches the true model. Warner and Guasch (2016) 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 (Alkhalifah and Choi, 2014) as demonstrated by Sun and Alkhalifah (2018c). In principle, the misfit functions designed for all these approaches try to measure the departure of the resulting matching filter from a Dirac delta function, e.g., with a penalty or a normalization term. Sun and Alkhalifah (2018a) reanalyze such comparison based on the optimal transport theory. They explain that due to the oscillatory nature of the seismic data signal, the original seismic data is not a probability distribution in nature. Per requirement of the optimal transport, any modification to the seismic data to make them a distribution would either alter its phase or amplitude, thus, make the resulting inversion process less accurate and unstable. Instead, Sun and Alkhalifah (2018a, 2019) proposed to compare the resulting matching filter with the Dirac delta function based on the optimal transport theory. A modification to the matching filter which makes it become a distribution would not alter the original seismic data signal, and this makes the inversion more stable and accurate. Based on the solid mathematical foundation of optimal transport theory, they show that the penalty term is embedded in the expression of the Wasserstein metric W_2 and the normalization term in AWI comes naturally per the requirement of a probability distribution, i.e., the nonnegative and mass conservation condition.

In this paper, we propose a method, designated as 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 di-

rectly would focus the inversion on the dominant events with high amplitude, and this makes the WTI method admit low-resolution results and limits its application. To resolve this issue, we propose to apply the cross-correlation between the penalty function t^2 and the resulting matching filter computed from the measured and predicted data. The resulting traveltimes shift, which minimizes the cross-correlation, would indicate the defocusing of the matching filter. The least square errors of the resulting traveltimes shift can be formulated as a misfit function. We will generalize our result and show that the traveltimes shift in ATI has a clear physical meaning and it is the first order moment, 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. Since the moment of a distribution is based on statistical assumption, other statistical measures, like the geometric mean and the median, can also be used as misfit functions.

In the following, we will start with a review of the WTI method at first and then present the theory of ATI. After that, we compare the proposed ATI method with the WTI and AWI methods. We use a cross-well Gaussian anomaly, modified Marmousi model and a real data from the Gulf of Mexico to demonstrate the properties and effectiveness of the ATI method.

REVIEW OF WAVE-EQUATION TRAVELTIME INVERSION

The method presented here is partially inspired by the wave-equation traveltimes inversion (WTI). So we start with a brief review of the method and for details, please look into Luo and Schuster (1991).

WTI tries to minimize the misfit function defined by the square errors of the traveltime shift $\Delta\tau$:

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

In order to simplify the notation here, we omit the summation over sources and receivers. 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.,

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

where $d(t)$ is the measured data and $p(t)$ is the predicted data computed using the current velocity model. As $\Delta\tau$ corresponds to the maximum value of $f_{\text{WTI}}(\tau)$, we would have

$$\frac{df_{\text{WTI}}(\tau)}{d\tau}|_{\tau=\Delta\tau} = \int d'(t + \Delta\tau)p(t)dt = 0, \quad (3)$$

where $d'(t) = \frac{dd(t)}{dt}$ denotes the time derivative. Equation 3 is the connective function, which relates the traveltime shift $\Delta\tau$ to the predicted data $p(t)$ and this relation can be used to compute the adjoint source of misfit function of equation 1. Here, we present the adjoint source for WTI directly (for detail derivation, please refer to Luo and Schuster (1991)):

$$\delta s_{\text{WTI}} = \frac{d'(t + \Delta\tau)\Delta\tau}{E}, \quad (4)$$

where E is defined as

$$E = \int d'(t + \Delta\tau)p(t)dt. \quad (5)$$

ADAPTIVE TRAVELTIME INVERSION

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

$$J_{\text{ATI}} = \frac{1}{2} \Delta\tau^2. \quad (6)$$

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 measured data to the predicted data :

$$d(t) * w(t) = p(t), \quad (7)$$

where $*$ denotes the convolution operation. Equation 7 is a linear equation, and the matching filter $w(t)$ can be computed either in the time domain, or more straightforward in frequency domain as follows:

$$w(t) = \mathcal{F}^{-1} \left[\frac{\mathcal{F}[p(t)] \overline{\mathcal{F}[d(t)]}}{\mathcal{F}[d(t)] \overline{\mathcal{F}[d(t)]} + \epsilon} \right], \quad (8)$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier transform and its inverse, respectively. Overline denotes the complex conjugation operation, and ϵ is a small positive number to avoid dividing over zero.

If the velocity model for producing the predicted data $p(t)$ is correct. The resulting matching filter should end up being an (approximated) Dirac delta function. Otherwise, its coefficients would spread out over non-zero lags. Similarly, we define a cross-correlation function $f_{\text{ATI}}(\tau)$ to measure such defocusing:

$$f_{\text{ATI}}(\tau) = \frac{1}{2} \int (t + \tau)^2 w^2(t) dt. \quad (9)$$

The time shift $\Delta\tau$ defined in the ATI misfit function of equation 6 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 (approximated) Dirac delta function, with the minimum value of $f_{\text{ATI}}(\tau)$ corresponding to time shift $\Delta\tau = 0$. Otherwise, the time shift $\Delta\tau$ is nonzero and by minimizing the square errors of equation 6, we can update the velocity model.

Similarly, as $\Delta\tau$ corresponds to the minimum value of $f_{\text{ATI}}(\tau)$, its first-order derivative at the minimum would be zero. Thus, we use the corresponding connective function given

by:

$$\frac{df_{\text{ATI}}(\tau)}{d\tau}|_{\tau=\Delta\tau} = \int (t + \Delta\tau)w^2(t)dt = 0. \quad (10)$$

For WTI, the connective equation of 3 implicitly defines the relation between the time shift $\Delta\tau$ and the predicted data $p(t)$. For ATI, the time shift $\Delta\tau$ can be resolved from equation 10 directly and it shows an explicit relation to the predicted data $p(t)$ via the matching filter $w(t)$ (note the explicit relation between $w(t)$ and $p(t)$ defined by equation 8).

By a simple arithmetic operation, we can express the time shift $\Delta\tau$ in equation 10 as :

$$\Delta\tau = -\frac{\int tw^2(t)dt}{\int w(t)^2dt} = -\frac{\int tw^2(t)dt}{||w||_2^2}. \quad (11)$$

Substituting the expression of the time shift $\Delta\tau$ in equation 11 into equation 6, we have the final explicit formula for ATI misfit function:

$$J_{\text{ATI}} = \frac{1}{2} \left[\frac{\int tw^2(t)dt}{||w||_2^2} \right]^2. \quad (12)$$

In order to update the velocity model, we need to compute the adjoint source, which is the derivative of the misfit function with respect to the predicted data:

$$\delta s_{\text{ATI}} = \frac{\partial w}{\partial p} \frac{\partial \Delta\tau}{\partial w} \frac{\partial J_{\text{ATI}}}{\partial \Delta\tau}, \quad (13)$$

where $\partial J_{\text{ATI}}/\partial \Delta\tau$ can be easily computed using equation 6,

$$\frac{\partial J_{\text{ATI}}}{\partial \Delta\tau} = \Delta\tau, \quad (14)$$

and $\partial \Delta\tau/\partial w$ can be computed from equation 11:

$$\frac{\partial \Delta\tau}{\partial w} = -\frac{2(t + \Delta\tau)w(t)}{||w||_2^2}. \quad (15)$$

The substitution of equations 14 and 15 into equation 13 leads to the adjoint source:

$$\delta s_{\text{ATI}} = -\frac{\partial w}{\partial p} \frac{2(t\Delta\tau + \Delta\tau^2)w(t)}{||w||_2^2}. \quad (16)$$

The left term $\partial w / \partial p$ is the derivative of the matching filter with respect to the predicted data, and it can be computed using the adjoint analysis of equation 8:

$$\frac{\partial w}{\partial p} = \mathcal{F}^{-1} \text{diag} \left[\frac{\mathcal{F}[d(t)]}{\mathcal{F}[d(t)]\overline{\mathcal{F}[d(t)]} + \epsilon} \right] \mathcal{F}. \quad (17)$$

COMPARISON OF WTI, ATI, AWI

In this section, we will outline the theoretical relation that ATI has with WTI and AWI, and show that the method allows for a generalization that could have AWI as a special case. We will also discuss some challenging models that ATI or AWI seemingly would fail.

ATI and WTI

Comparing ATI with WTI, the main difference is the cross-correlation functions of equations 2 and 9. In WTI, the cross-correlation function of equation 2 correlates the measured data and predicted data directly and this kind of direct cross-correlation between waveforms makes it focus on the events with large amplitude, and thus, prevents WTI from handling properly scattered energy in the data as such scattered energy normally has small amplitude. As reported by Luo and Schuster (1991), WTI has relatively low resolution compared to full waveform inversion (Zhou et al., 1995, 1997; Zheng et al., 2013).

For ATI, the cross-correlation strategy is a bit different: a matching filter $w(t)$ is computed first using equation 7. The matching filter provides a global comparison between the predicted data and the measured data and this feature makes the resulting ATI method can handle the "cycle skipping" and thus makes it more "adaptive." The ATI cross-correlation of equation 9 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

9 of ATI does not cross-correlate the predicted and measured data directly, 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 a high-resolution full-waveform inversion method. We include "traveltime" in the name to highlight its similarity with the wave-equation traveltime inversion method.

ATI and AWI

In adaptive waveform inversion (AWI) (Warner and Guasch, 2016), if we choose the penalty function as $T^2(t) = t^2$, the misfit function of AWI is given by:

$$J_{\text{AWI}} = \frac{1}{2} \frac{\int t^2 w^2(t) dt}{\|w\|_2^2}. \tag{18}$$

If we compare equation 18 with equation 12, we can see that they are generally similar. We can express them under a more general form based on the n-th order moment,

$$\mu_n = \int t^n q(t) dt, \tag{19}$$

where $q(t)$ is a distribution computed from the matching filter:

$$q(t) = \frac{w^2(t)}{\int w^2(t) dt}. \tag{20}$$

Thus, we can define a generalized misfit function as:

$$J_n = \mu_n^m, \tag{21}$$

where the power m is introduced to guarantee the misfit function of J_n is nonnegative: if the order of the moment n is odd, m equals to two otherwise it is set to be one. With those

generalization, in equation 19, for $n = 1$, the resulting first order moment μ_1 is the mean and it is corresponding to travelttime shift $\Delta\tau$ of equation 11 for ATI (ignore the negative sign). For $n = 2$, the resulting second order moment μ_2 corresponds to the misfit function of equation 18 for AWI.

As we explained before, if the velocity model is correct, the resulting distribution computed from the matching filter using equation 20 is an (approximated) Dirac delta function. Normally, we can approximate a Dirac delta function by a Gaussian function $g(t)$ with its mean value as c and variance as σ^2

$$g(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-c)^2}{2\sigma^2}}. \quad (22)$$

As for the Dirac delta function, both the mean value c , which is the first order moment and corresponds to the travelttime shift $\Delta\tau$ in ATI and the variance σ^2 , which equals to the second order moment (supposing the mean is already zero) and corresponds to the misfit function of AWI, should be zeros. Thus, minimizing the misfit function for both ATI and AWI would push energy in the matching filter towards to zero lag resulting in a Dirac delta function. The difference is that for ATI, it forces its mean to zero value, while for AWI, it tries to reduce the variance (supposing the mean is already zero.)

In summary, 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. 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 may highlight additional features.

Challenging models

We can design simple challenging models for which ATI seemingly would fail. However, some of them would hardly happen during practical inversions of real data. We discuss some of these related issues here.

Suppose the predicted and measured data differ only by a constant time shift. In this case, the resulting matching filter appears to have only one nonzero coefficient at a nonzero lag. If we substitute such matching filter into equation 11 and 16 , we would have a zero-value adjoint source. Thus, the gradient would be zero, and it seems that we can't update the model in this situation. However, the seismic data are band-limited, and in the computation of the matching filter, we also introduce ϵ in equation 8 for stabilization considerations. Those would make the resulting matching filter band-limited as well and hardly only non zero at a single time lag. This would consequently lead to a finite adjoint source and gradient for updating the velocity model. In other words, a pure constant time shift between the predicted and measured data indicates that the current model is a constant scale of the true model and this would hardly happen in reality as well.

The second challenging possibility is that the computation of the traveltime shift using equation 10 ignores the polarity of the matching filter (note the square over the matching filter $w(t)$). We might assume that ATI will recover a model for which the polarity of the predicted data was reversed. For AWI misfit function of equation 18, those can be an issue as well. However, as pointed out by Warner and Guasch (2016): "In any real-world data set, there will always be at least one reference arrival that is impossible, or at least extremely difficult, to reverse using any reasonable physical model. These include the direct arrival, the primary sea-bottom reflection, and most primary refractions, head waves, and

turning arrivals. Provided that at least one of these reference arrivals is present in the field data, then the global minimum will always correspond to a model that also has the correct polarity for all other events.” In short, due to the redundancy of the seismic dataset, it is extremely hard to create a model that fully reverse all traces at all offsets and shots while still in the region of a global minimum.

Another challenging possibility is that the measured data are dominated with multiples such as free surface or internal multiples. This would increase the nonlinearity of the dataset as well as the resulting misfit function. The computed matching filter, in this case, faces the prospect of matching multiples to primaries or vice versa. The resulting adjoint source would be less reliable in capturing the error in the velocity model. The convergence of the inversion becomes slow and may still get stuck in a local minimum if the initial model is relatively far away from the true one. In this sense, the matching filter approach can reduce the ”cycle-skipping” but can not mitigate it completely, a successful application of the matching-filter approach still favors a relatively good initial velocity model.

EXAMPLES

In this section, we will first analyze the convexity of the various objective functions. Then we test the approaches on a simple Gaussian anomaly, followed by the Marmousi model. Finally, we apply the proposed ATI method to a marine field data from the Gulf of Mexico.

Convexity of the misfit functions

In the first example, we examine the convexity of the misfit functions based on the L2 norm, WTI, ATI, and AWI. In Figure 1a, the black line is the measured data $d(t) = f(t)$,

while the blue line is the predicted data $p(t + \tau) = f(t + \tau)$. Here $f(t)$ is given by a Ricker wavelet with a peak frequency of 10 Hz. Thus, by controlling the value of τ , we can time shift the predicted data $p(t + \tau)$ relative to the measured data $d(t)$. Figure 1b shows the misfit function values changes respect to the time shift τ . It is obvious that misfit function for L2 norm has a local minimum, while WTI, ATI and AWI misfit functions are all convex and free of the local minimum. The misfit values for all three are almost the same. For WTI, it measures the square errors of the time shift, so its misfit function is proportional to τ^2 . In our setup, the resulting matching filter can be approximated by a shifted Dirac Delta function $\delta(t - \tau)$. If we substitute it into the ATI and AWI misfit function of equation 12 and 18 respectively, we can realize that the misfit-function values of ATI and AWI are also proportional to τ^2 . In Figure 1c, we take a close look at the area of small time shift. We can see that the AWI misfit function has a relatively larger value compared to other misfit functions. It is due to the band-limited nature of the signal. The resulting matching filter would never be an exact delta function with only none zero values at $t = 0$. AWI measures the variance of such approximate delta functions, and thus, it would have higher value. However, ATI measures the mean, which tends to be lower.

Cross-well Gaussian anomaly

The second example is a cross-well Gaussian anomaly model. As shown in Figure 2a, the model size is about 8 km in depth and 2 km laterally. We setup 400 shots and 400 receivers with the same interval spacing of 20 m on each side. We use a constant velocity of 2500 m/s as the initial model, which is shown in Figure 2b. We plot a common shot record for a shot located at depth 4 km in Figure 2c. The red traces are for the true model while the black ones are for the initial model. It is easy to see that the record corresponding to the

initial model is "cycle skipped," especially at the far offset. We invert the full band data directly without any frequency continuation. The inversion is performed using a nonlinear conjugate gradient method. A backtracking line search strategy is used with maximum line search step 30 m/s and a minimum of 5 m/s. We run for a maximum of 40 iterations. Figure 3 shows the inverted models using different misfit functions. We can see that the result of Figure 2a using the L2-norm misfit function failed to capture the two Gaussian anomalies correctly. Meanwhile, other misfit functions recover the two Gaussian anomalies correctly. Figure 4 shows the vertical profile at location 1 km; it further demonstrates the failure of the L2 norm misfit function and the success of WTI, ATI and AWI misfit functions. Figure 5 shows common shot records using the inverted models. The red traces are for the measured record, while the black traces are the record corresponding to the inverted models. Figure 5a indicates that because of "cycle skipping," the L2 norm can't recover the model correctly, which leads to a poor fit of the data as well. However, the other three misfit functions: WTI, ATI, and AWI can resolve the cycle skipping issues; the records from inverted models fit the measured records pretty well. At last, in Figure 6, we show the gradients of the first iteration for different misfit functions. Figure 6a is computed by the difference between the initial and true velocities. After normalization, it can be considered as the reference for a "true" gradient. Figure 6b is the gradient for the L2 norm misfit function, it is cycle-skipped and the sign for the low-velocity perturbation in the deeper part is wrong. The gradients for other misfit functions show the correct sign for updating and have similar shapes as well. WTI method performed well in this example because the majority of the information was provided by the dominant transmission arrivals. With a more complex model, the story would be different.

Modified Marmousi Model

The second example corresponds to the more complex modified Marmousi model. Unlike previous cross-well example where only transmission waves are under consideration, here the waves are complex as it includes various wave modes like diving, reflection and refraction waves. As shown in Figure 7a, the true velocity v_{true} model extends 2 km in depth and 8 km laterally. The initial velocity $v_{\text{ini}} = v(z)$ is shown in Figure 7b. 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. 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. Figure 8 shows a comparison of the records computed from the true model (left panel) and the initial model (right panel) for a source at location 400 m. It is obvious that it is cycle skipped at far offsets for the current frequency band.

In these time-domain inversions, we apply 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 a frequency range of 3 Hz to 10 Hz and no window muting or damping is applied to the data. This is a challenge for the conventional L2 misfit function as for this frequency band we face the "cycle skipping" issue. We run the inversion over 200 iterations using a nonlinear conjugate gradient method. In each iteration, we normalize the gradient by its maximum value and perform a backtracking line search with a maximum step of 100 m/s. We allow for a maximum of 5 iterations for the line search, and the minimum step for line search is 15 m/s. The inverted result for the conventional L2-norm misfit function is shown in Figure 9a. 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 9b 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 9c and 9d shows results from ATI and AWI methods. We can see that both methods recover the model well, including 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. We compare the resulting shot records for a source located at 400 m for the WTI (Figure 10), ATI (Figure 11), and AWI(Figure 12) misfit functions, respectively. We plot the record corresponding to the true model in the left panel for reference and the record for the inverted models on the right panel. We can see that the records for ATI and AWI fit the true model well. The high accuracy of the inverted model is well demonstrated with the good alignment of events between the true record and inverted ones, especially at far offsets. Considering the initial velocity is far away from the true one, the proposed method reduced the "cycle skipping" and converged towards the global minimum successfully. WTI, in general, performed worse than the ATI and AWI methods. The weakness of WTI in inverting scattering events is especially evident in this example (note the yellow ellipse in Figure 10). The fitting at far offset is not good as well compared with ATI and AWI result.(Note the white ellipse in Figure 10).

Figure 13a shows the misfit functions over iterations. We can see that the misfit for the L2 norm drops very quickly, however, gets stuck at a relatively large value while other misfit functions drop to relatively low values gradually. AWI misfit function (the black line) shows relatively large residuals due to the band-limited nature of the data. Figure 13b shows the least square misfit of the residuals over iterations for different methods. Note as the least squares of the residuals are not consistent with the definition of the misfit functions of WTI,

ATI, and AWI, oscillation can be observed. Due to the global measurement behavior of ATI and AWI, at the initial stage, the least squares of the residuals would increase. This can be considered as an adaptation phase. Afterward, these global comparison approaches will gradually reduce the residuals and converge to a model with smaller residuals compared to the L2-norm method. From the displayed curves, we can also observe that the ATI method shows better accuracy than WTI and AWI method.

Figure 14 shows the inversion evolution of the WTI misfit function approach. We can see that the WTI method favors updating the low wavenumber components of the model and can not recover the fine scale structure, especially in the middle deeper part. On the other hand, the ATI and AWI methods, shown in figures 15 and 16, can update the low and high wavenumber components of the model simultaneously, and they show similar features in recovering the area where the data have better coverage, with a flow that goes from inwards to outwards, shallow to deep, and gradually updates the model.

At last, we show the evolution of the matching filter distribution for a shot at distance 2 km. For ATI method, it is shown in Figure 17 while AWI method in Figure 18. The matching filter distributions are computed according to equation 20, and the extra normalization term included helps with the demonstration and examination of its focusing properties free of amplitude effects. By the figures, it is obvious to see that the resulting matching filter for both ATI and AWI method becomes more focusing during iterations and finally converges to a band-limited Dirac delta function.

Real Data Examples

In this last example, we try the proposed ATI method on a marine data from the Gulf of Mexico and compare it with the conventional L2-norm misfit function approach. The dataset has 1000 shots with shot point interval of 25 m and a receiver group interval of 25 m. Each shot record was recorded on 120 channels for 6 seconds. The water depth for this region is around 350 m. The dataset has strong multiples as shown in Figure 19a. The maximum offset is about 3.2 km. This limited offset is a challenge to the conventional L2 norm misfit function as such small offsets would result in diving waves with very limited depth penetration. We want to show in this example that our proposed misfit function can also use the scattered energy in the inversion even with a poor starting model. The minimum usable frequency is about 6 Hz. Figure 19b shows the same common shot record of Figure 19a after low-pass filtering with maximum frequency equal to 6Hz. In the inversion, we also mute frequencies below 3 Hz in the data and start our inversion with a low-pass filter applied to the data with maximum frequency ranging from 6 Hz to 10 Hz every 1 Hz, 10 Hz to 16 Hz every 2 Hz, and 16 Hz to 25 Hz every 5 Hz. For the ATI inversion method, we will invert for 20Hz to 25Hz with L2-norm misfit function to further reduce the artifacts and improve the results. In the inversion, we invert for the source wavelet for each shot using the traces at near offset. Figure 20 shows one typical inverted source wavelet and its spectrum in the frequency band 3-20 Hz. We were provided with a far-field source wavelet. We tried using this source wavelet in inversion after proper processing, such as deghosting. However, our tests show that inverting the source using direct arrivals provides the most suitable source wavelet. This indicates that including the water bottom reflectivity in the wavelet is important for a successful inversion.

Figure 21a shows the initial velocity used for inversion. Figure 21b shows the inverted model using the L2 norm. Figure 21c shows the inverted model from the proposed ATI method and Figure 21d is the final result after a successive inversion using L2-norm refinement of ATI output of figure 21c. We can see that the result in Figure 21b corresponding to the L2 norm has relatively strong artifacts compared to the final result in Figure 21d.

We select one shot at the middle of the model to check whether the resulting models produce data that can fit the measured data. Figure 22 shows the record comparison of the measured data with the predicted data using initial model (note here and below, in the figure, red traces correspond to the measured record while black traces corresponds to the predicted one). Figure 23 shows the comparison of the records of the measured data with the predicted one modeled using the inverted model by the L2 norm method. Meanwhile, Figure 24 compares the measured record with the predicted one obtained using the final inverted model based on the proposed method(ATI plus L2-norm.) We can see that both the L2 norm and the proposed method show better fitting at far offsets than the initial model. Note in the area designated by the blue rectangles which are zoomed in and shown in Figure 25a and Figure 25b: the record using the proposed method shows better fitting to the scattered energy (reflections) than L2 norm. This indicates improved accuracy of the proposed method and its potential to use reflections for velocity updating.

In order to further evaluate the accuracy of our inverted model, we perform imaging using a wave-equation migration method, as well as, output angle domain common image gathers. In Figure 26, 27 and 28, we show the images corresponding to the initial velocity model, the inverted model using the L2 norm misfit function and the final inverted result using our proposed method, respectively. We can see that after FWI updating both images of the L2 norm (Figure 27) and the proposed method (Figure 28) are better focused than

the image using the initial model (Figure 26). However, the image of L2 norm (Figure 27) has uneven layers due to the artifacts in the model, while the image for our proposed method (Figure 28) shows more consistent geological structures, e.g., the area noted by yellow arrows. We display corresponding angle-domain common image gathers (CIGs) every 1.1km in Figure 29, 30 and 31 (We processed the output gather to reduce the effect of multiples using Radon transform.) It is evident that the CIGs (Figure 31) using the velocity model from our proposed method improve the flatness of the gathers considerably compared to the CIGs using the initial model (Figure 29). The result from the model of the L2-norm (Figure 30) still has relatively large residual moveout.

DISCUSSION

The main purpose of this paper is to demonstrate the theoretical relationship between WTI, ATI and AWI. We attempt to understand how the matching filter evolves during the inversion process. We can obtain a relatively robust ATI misfit function by minimization of the mean value of the resulting matching filter distribution only. We may suspect that the inversion result provides eventually a matching filter with zero mean and a considerable variance. However, according to both the simple Gaussian lens and the relatively complicated Marmousi examples, this seems not be the case. Since the resulting matching filter is rather complex, the distribution of the matching filter is reasonably random. When the velocity model is inaccurate, the matching filter distribution often has a non-zero mean. When we update the model, the mean value of the matching filter distribution reduces and so does its variance. As long as the velocity is inaccurate, the resulting matching filter distribution is asymmetric with a non-zero mean. The inversion process as we saw will not try to reduce the mean value by increasing the variance. However, adding an extra term, which

controls the compactness (variance) of the matching filter distribution, may accelerate the convergence of the inversion of the ATI method. For example, the entropy (Machta, 1999) of a distribution can be an ideal candidate:

$$\text{Entropy}(q) = - \int q(t) \log [q(t)] dt. \tag{23}$$

The entropy of a distribution has its maximum when the distribution is uniformly distributed and decreases its value when the distribution becomes more compactly focused. Thus, we can add the negative of the entropy into the ATI misfit function as a regularization term. This extra regularization term is expected to reduce the variance of the resulting matching filter and may improve the convergence of the inversion. In this paper, we focus on the ATI formula itself and do not apply the extra entropy regularization term in the examples.

Alternatively, AWI corresponds to the second-order moment of the matching filter, thus improvements may be achieved by considering higher-order moments such as the skewness (third order moment) or the kurtosis (fourth order moment). We will explore all those interesting possibilities in the near future.

CONCLUSION

We proposed a robust misfit function for full waveform inversion, we refer to as adaptive traveltime inversion (ATI). The inversion is formulated by minimizing the least square errors of the traveltime shift computed from cross-correlation of a penalty function with matching filter. The matching filter is computed by a deconvolution of the measured and predicted

data. Compared to conventional L2-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 traveltimes shift is not computed using cross-correlation of the measured and predicted data directly, ATI admits high accuracy and resolution than the WTI method. It also allows us to extend the misfit function to a more generalized form regarding 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. The misfit function of adaptive waveform inversion (AWI) can be considered as a special case of the second-order moment.

ACKNOWLEDGMENTS

We thank the associate editor F. Liu and the reviewers C. Chen and P. Williamson for their constructive comments that helped to improve the paper. The authors also thank the SWAG group for useful discussion and the resources of the shaheen supercomputing laboratory at KAUST.

REFERENCES

Alkhalifah, T., and Y. Choi, 2014, From tomography to full-waveform inversion with a single objective function: *Geophysics*, **79**, no. 2, R55–R61.

Alkhalifah, T., B. Sun, Y. Choi, F. Alonaizi, and M. Almalkl, 2018, Sparse frequencies data inversion: an application to a near surface experiment: 80th EAGE Conference and Exhibition.

Biondi, B., and W. W. Symes, 2004, Angle domain common image gathers for migration velocity analysis by wavefield continuation imaging: *Geophysics*, **69**, no. 5, 1283–1298.

Datta, D., and M. K. Sen, 2016, Estimating a starting model for full-waveform inversion using a global optimization method: *Geophysics*, **81**, no. 4, R211–R223.

Debens, H. A., F. Mancini, M. Warner, and L. Guasch, 2017, Full-bandwidth adaptive waveform inversion at the reservoir: *SEG Technical Program Expanded Abstracts*, 1378–1382.

Engquist, B., and B. Froese, 2014, Application of the wasserstein metric to seismic signals: *Communications in Mathematical Sciences*, **9**, no. 1, 79–88.

Engquist, B., B. D. Froese, and Y. Y., 2016, Optimal transport for seismic full waveform inversion: *Communications in Mathematical Sciences*, **14**, no. 2, 79–88.

Jin, S., and R. Madariaga, 1993, Background velocity inversion with a genetic algorithm: *Geophysical Research Letters*, **20**, 93–96.

———, 1994, Nonlinear velocity inversion by a two step monte carlo method.: *Geophysics*, **59**, no. 4, 577–590.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi, 1983, Optimization by simulated annealing: *Science*, **220**, no. 4598, 671–680.

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, Philadelphia, Expanded Abstracts, 206–220.
- Luo, S., and P. Sava, 2011, A deconvolution based objective function for wave equation inversion: SEG Technical Program Expanded Abstracts, 2788–2792.
- Luo, Y., and G. T. Schuster, 1991, Wave equation travelttime inversion: Geophysics, **56**, no. 5, 645–653.
- Machta, J., 1999, Entropy, information, and computation: American Journal of Physics, **67**, no. 12, 1074–1077.
- Métivier, L., R. Brossier, Q. Mrigot, E. Oudet, and J. Virieux, 2016, Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion: Geophysical Journal International, **205**, no. 1, 345–377.
- Operto, S., J. Virieux, J.-X. Dessa, and G. Pascal, 2006, Crustal seismic imaging from multifold ocean bottom seismometer data by frequency domain full waveform tomography: Application to the eastern nankai trough: Journal of Geophysical Research: Solid Earth, **111**, B09306.
- Plessix, R. E., and C. Perkins, 2010, Full waveform inversion of a deep water ocean bottom seismometer dataset: First Break, **28**, 71–78.
- Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, **64**, no. 3, 888–901.
- Qiu, L., J. Ramos-Martnez, A. Valenciano, Y. Yang, and B. Engquist, 2017, Full-waveform inversion with an exponentially encoded optimal-transport norm: SEG Technical Program Expanded Abstracts, 1286–1290.
- Sambridge, M., and K. Mosegaard, 2002, Monte carlo methods in geophysical inverse problems.: Reviews of Geophysics, **40**, no. 3, 1009.

- Sansigre, A. M., and A. Ratcliffe, 2014, A probabilistic qc for cycle-skipping in full waveform inversion: SEG Technical Program Expanded Abstracts 2014, 1105–1109.
- Sava, P., and S. Fomel, 2006, Time-shift imaging condition in seismic migration: *Geophysics*, **71**, no. 6, S209–S217.
- Sen, M. K., and P. L. Stoffa, 1992, Rapid sampling of model space using genetic algorithms: examples from seismic waveform inversion: *Geophysical Journal International*, **108**, no. 1, 281–292.
- Shen, P., and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: *Geophysics*, **73**, no. 5, VE49–VE59.
- Shen, X., I. Ahmed, A. Brenders, J. Dellinger, J. Etgen, and S. Michell, 2018, Full-waveform inversion: The next leap forward in subsalt imaging: *The Leading Edge*, **37**, no. 1, 67b1–67b6.
- Shin, C., and Y. Cha, 2008, Waveform inversion in the laplace domain: *Geophysical Journal International*, **173**, no. 3, 922–931.
- , 2009, Waveform inversion in the laplace fourier domain: *Geophysical Journal International*, **177**, no. 3, 1067–1079.
- Sirgue, L., O. I. Barkved, J. Dellinger, U. A. J. Etgen, and J. H. Kommedal, 2010, Full waveform inversion: The next leap forward in imaging at valhall: *First Break*, **28**, 65–70.
- Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, no. 1, 231–248.
- 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.
- , 2018b, Automatic wave-equation migration velocity analysis by focusing subsurface

- virtual sources: *Geophysics*, **83**, no. 2, U1–U8.
- , 2018c, Mitigate cycle skipping in fwi: A generalized instantaneous travel-time approach: 80th EAGE Conference and Exhibition.
- , 2019, A robust waveform inversion using a global comparison of modeled and observed data: *The Leading Edge*, **38**, no. 3, 185–192.
- Symes, W. W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: *Geophysics*, **56**, no. 5, 654–663.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, no. 8, 1259–1266.
- Van Leeuwen, T., and W. Mulder, 2008, Velocity analysis based on data correlation: *Geophysical Prospecting*, **56**, no. 6, 791–803.
- Van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation travelttime tomography: *Geophysical Journal International*, **182**, no. 3, 1383–1394.
- Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application using multicomponent measurements of seismic data collection: *Geophysics*, **79**, no. 2, R63–R77.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26.
- Wang, B., Y. Kim, C. Mason, and X. Zeng, 2008, Advances in velocity model-building technology for subsalt imaging: *Geophysics*, **73**, no. 5, VE173–VE181.
- Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: *Geophysics*, **81**, no. 6, R429–R445.
- Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. tekl,

L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3d full-waveform inversion: *Geophysics*, **78**, no. 2, R59–R80.

Woodward, M. J., D. Nichols, O. Zdraveva, P. Whitfield, and T. Johns, 2008, A decade of tomography: *Geophysics*, **73**, no. 5, VE5–VE11.

Yang, Y., and B. Engquist, 2018, Analysis of optimal transport and related misfit functions in full-waveform inversion: *Geophysics*, **83**, no. 1, A7–A12.

Yang, Y., B. Engquist, J. Sun, and B. F. Hamfeldt, 2018, Application of optimal transport and the quadratic wasserstein metric to full-waveform inversion: *Geophysics*, **83**, no. 1, R43–R62.

Zhang, Z., T. Alkhalifah, E. Z. Naeini, and B. Sun, 2018, Multiparameter elastic full waveform inversion with facies-based constraints: *Geophysical Journal International*, **213**, no. 3, 2112–2127.

Zheng, Y., Y. Wang, and X. Chang, 2013, Wave equation travelttime inversion: Comparison of three numerical optimization methods: *Computers and Geosciences*, **60**, 88–97.

Zhou, C., W. Cai, Y. Luo, G. T. Schuster, and S. Hassanzadeh, 1995, Acoustic wave equation travelttime and waveform inversion of crosshole seismic data: *Geophysics*, **60**, 765–773.

Zhou, C., G. T. Schuster, S. Hassanzadeh, and J. M. Harris, 1997, Elastic wave equation travelttime and waveform inversion of crosswell data: *Geophysics*, **62**, no. 3, 853–868.

LIST OF FIGURES

- 1 a) The measured and predicted data; b) The corresponding misfit functions; c) A
zoom view of the figure b).
- 2 a) The true velocity; b) The initial velocity; c) Common shot records (red corre-
sponding to the true model, black for the initial model).
- 3 The inverted Gaussian anomaly model using a) L2; b) WTI; c) ATI; d) AWI misfit
functions, respectively.
- 4 Vertical profiles for the inverted model using different misfit functions.
- 5 Comparisons between the measured common shot record(red) and shot records
computed using the inverted models(black) for a) L2 ; b) WTI; c) ATI; d) AWI approaches.
- 6 a) The normalized velocity perturbation $\delta v = (v_{\text{ini}} - v_{\text{true}})/\max(v_{\text{ini}} - v_{\text{true}})$; Gra-
dients from the first iteration of b) L2 norm; c) WTI; d) ATI; e) AWI approaches.
- 7 a) The true velocity and b) the initial velocity for the modified Marmousi model
example.
- 8 A comparison of common shot records corresponding to the true model(left panel)
and the initial model(right panel).
- 9 The inverted velocity using a) L2 norm ; b) WTI; c) ATI ; d) AWI misfit functions.
- 10 A comparison of the common shot record for the true model(left panel) and the
inverted model using WTI misfit function(right panel).
- 11 A comparison of the common shot record for the true model(left panel) and the
inverted model using ATI misfit function(right panel).
- 12 A comparison of the common shot record for the true model(left panel) and the
inverted model using AWI misfit function(right panel).
- 13 a) The misfit function and b) the least square misfit functions of residual over iter-

ations for different methods. Note in figure (a), the misfit-function values are normalized by the maximum misfit-function values of each method while in figure (b), the misfit-function values are normalized by the maximum value of the L2-norm method.

14 Inverted models using the WTI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

15 Inverted models using the ATI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

16 Inverted models using the AWI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

17 The matching filter distribution for ATI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

18 The matching filter distribution for AWI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

19 a) The raw record of the real data; b) low pass filtering of a) with cutting frequency equals to 6Hz.

20 a) Inverted source wavelet in the frequency band 3-20Hz and b) its spectrum.

21 a) Initial velocity; inverted velocity by b) L2 norm and c) ATI misfit function d) final result after L2 norm inversion refinement of c).

22 Comparision of the common shot record for the measured data(red) and modeled data by initial model(black).

23 Comparision of the common shot record for the measured data(red) and modeled data by result inverted by L2 norm(black).

24 Comparision of the common shot record for the measured data(red) and modeled data by result inverted by ATI plus L2 norm(black).

25 Zoom-in view of the area denoted by blue rectangles in Figure 23 and Figure 24:
26
27 comparison of the common shot record for the measured data(red) and modeled data by
28 result inverted by a) L2 norm (black) b) ATI plus L2 norm(black).

29 Image by initial velocity (Figure 21a)

30 Image by inverted velocity using L2-norm misfit function(Figure 21b).

31 Image by inverted velocity using ATI plus L2-norm misfit function(Figure 21d).

32 Angle domain common image gathers for initial velocity shown(Figure 21a).

33 Angle domain common image gathers for inverted velocity by L2-norm misfit func-
34 tion(Figure 21b).

35 Angle domain common image gathers for inverted velocity by ATI plus L2-norm
36 misfit function(Figure 21d).

Figure 1a a) The measured and predicted data; b) The corresponding misfit functions; c) A zoom view of the figure b).

1313x1360mm (72 x 72 DPI)

Figure 1b a) The measured and predicted data; b) The corresponding misfit functions; c) A zoom view of the figure b).

1298x1359mm (72 x 72 DPI)

Figure 1c a) The measured and predicted data; b) The corresponding misfit functions; c) A zoom view of the figure b).

1960x1960mm (72 x 72 DPI)

Figure 2a a) The true velocity; b) The initial velocity; c) Common shot records (red corresponding to the true model, black for the initial model).

1387x1388mm (72 x 72 DPI)

Figure 2b a) The true velocity; b) The initial velocity; c) Common shot records (red corresponding to the true model, black for the initial model).

1387x1388mm (72 x 72 DPI)

Figure 2c a) The true velocity; b) The initial velocity; c) Common shot records (red corresponding to the true model, black for the initial model).

1471x1471mm (72 x 72 DPI)

Figure 3a The inverted Gaussian anomaly model using a) L2; b) WTI; c) ATI; d) AWI misfit functions, respectively.

1387x1388mm (72 x 72 DPI)

Figure 3b The inverted Gaussian anomaly model using a) L2; b) WTI; c) ATI; d) AWI misfit functions, respectively.

1387x1388mm (72 x 72 DPI)

Figure 3c The inverted Gaussian anomaly model using a) L2; b) WTI; c) ATI; d) AWI misfit functions, respectively.

1387x1388mm (72 x 72 DPI)

Figure 3d The inverted Gaussian anomaly model using a) L2; b) WTI; c) ATI; d) AWI misfit functions, respectively.

1387x1388mm (72 x 72 DPI)

Figure 4 Vertical profiles for the inverted model using different misfit functions.

1197x1360mm (72 x 72 DPI)

Figure 5a Comparisons between the measured common shot record(red) and shot records computed using the inverted models(black) for a) L2 ; b) WTI; c) ATI; d) AWI approaches.

1471x1471mm (72 x 72 DPI)

Figure 5b Comparisons between the measured common shot record(red) and shot records computed using the inverted models(black) for a) L2 ; b) WTI; c) ATI; d) AWI approaches.

1471x1471mm (72 x 72 DPI)

Figure 5c Comparisons between the measured common shot record(red) and shot records computed using the inverted models(black) for a) L2 ; b) WTI; c) ATI; d) AWI approaches.

1471x1471mm (72 x 72 DPI)

Figure 5d Comparisons between the measured common shot record(red) and shot records computed using the inverted models(black) for a) L2 ; b) WTI; c) ATI; d) AWI approaches.

1471x1471mm (72 x 72 DPI)

Caption : Figure 6a a) The normalized velocity perturbation $\delta v = (v_{\text{ext}\{\text{ini}\}} - v_{\text{ext}\{\text{true}\}}) / \max(v_{\text{ext}\{\text{ini}\}} - v_{\text{ext}\{\text{true}\}})$; Gradients from the first iteration of b) L2 norm; c) WTI; d) ATI; e) AWI approaches.

1252x1266mm (72 x 72 DPI)

Caption : Figure 6b. a) The normalized velocity perturbation $\delta v = (v_{\text{ini}} - v_{\text{true}}) / \max(v_{\text{ini}} - v_{\text{true}})$; Gradients from the first iteration of b) L2 norm; c) WTI; d) ATI; e) AWI approaches.

1252x1266mm (72 x 72 DPI)

Caption : Figure 6c. a) The normalized velocity perturbation $\delta v = (v_{\text{ini}} - v_{\text{true}}) / \max(v_{\text{ini}} - v_{\text{true}})$; Gradients from the first iteration of b) L2 norm; c) WTI; d) ATI; e) AWI approaches.

1252x1266mm (72 x 72 DPI)

Caption : Figure 6d. a) The normalized velocity perturbation $\delta v = (v_{ini} - v_{true}) / \max(v_{ini} - v_{true})$; Gradients from the first iteration of b) L2 norm; c) WTI; d) ATI; e) AWI approaches.

1252x1266mm (72 x 72 DPI)

Caption : Figure 6e. a) The normalized velocity perturbation $\delta v = (v_{\text{ini}} - v_{\text{true}}) / \max(v_{\text{ini}} - v_{\text{true}})$; Gradients from the first iteration of b) L2 norm; c) WTI; d) ATI; e) AWI approaches.

1252x1266mm (72 x 72 DPI)

Figure 7a. a) The true velocity and b) the initial velocity for the modified Marmousi model example.

1411x745mm (72 x 72 DPI)

Figure 7b. a) The true velocity and b) the initial velocity for the modified Marmousi model example.

1411x745mm (72 x 72 DPI)

Figure 8 A comparison of common shot records corresponding to the true model(left panel) and the initial model(right panel).

1310x1309mm (72 x 72 DPI)

Figure 9a. The inverted velocity using a) L2 norm ; b) WTI; c) ATI ; d) AWI misfit functions.

1411x795mm (72 x 72 DPI)

Figure 9b. The inverted velocity using a) L2 norm ; b) WTI; c) ATI ; d) AWI misfit functions.

1411x795mm (72 x 72 DPI)

Figure 9d. The inverted velocity using a) L2 norm ; b) WTI; c) ATI ; d) AWI misfit functions.

1411x795mm (72 x 72 DPI)

Figure 10 A comparison of the common shot record for the true model(left panel) and the inverted model using WTI misfit function(right panel).

1310x1309mm (72 x 72 DPI)

Figure 11 A comparison of the common shot record for the true model(left panel) and the inverted model using ATI misfit function(right panel).

1310x1309mm (72 x 72 DPI)

Figure 12 A comparison of the common shot record for the true model(left panel) and the inverted model using AWI misfit function(right panel).

1310x1309mm (72 x 72 DPI)

Figure 13a. a) The misfit function and b) the least square misfit functions of residual over iterations for different methods. Note in figure (a), the misfit-function values are normalized by the maximum misfit-function values of each method while in figure (b), the misfit-function values are normalized by the maximum value of the L2-norm method.

1342x1116mm (72 x 72 DPI)

Figure 13b. a) The misfit function and b) the least square misfit functions of residual over iterations for different methods. Note in figure (a), the misfit-function values are normalized by the maximum misfit-function values of each method while in figure (b), the misfit-function values are normalized by the maximum value of the L2-norm method.

1331x1096mm (72 x 72 DPI)

Figure 14a. Inverted models using the WTI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 14b. Inverted models using the WTI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 14c. Inverted models using the WTI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 14d. Inverted models using the WTI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 14e. Inverted models using the WTI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 14f. Inverted models using the WTI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 15a. Inverted models using the ATI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 15b. Inverted models using the ATI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 15c. Inverted models using the ATI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 15d. Inverted models using the ATI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 15e. Inverted models using the ATI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 15f. Inverted models using the ATI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 16a. Inverted models using the AWI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 16b. Inverted models using the AWI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 16c. Inverted models using the AWI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 16d. Inverted models using the AWI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 16e. Inverted models using the AWI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 16f. Inverted models using the AWI misfit function at iteration a) 10; b) 20; c) 40; d) 50; e) 100; f) 200.

2187x1215mm (72 x 72 DPI)

Figure 17a. The matching filter distribution for ATI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 17b. The matching filter distribution for ATI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 17c. The matching filter distribution for ATI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 17d. The matching filter distribution for ATI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 17e. The matching filter distribution for ATI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 17f. The matching filter distribution for ATI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 18a. The matching filter distribution for AWI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 18b. The matching filter distribution for AWI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 18c. The matching filter distribution for AWI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 18d. The matching filter distribution for AWI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 18e. The matching filter distribution for AWI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 18f. The matching filter distribution for AWI misfit function at iteration: a) 10 ; b) 20 ; c) 40 ; d) 50 ; e) 100 ; f) 199.

2371x1351mm (72 x 72 DPI)

Figure 19a. a) The raw record of the real data; b) low pass filtering of a) with cutting frequency equals to 6Hz.

1240x1300mm (72 x 72 DPI)

Figure 19b. a) The raw record of the real data; b) low pass filtering of a) with cutting frequency equals to 6Hz.

1240x1300mm (72 x 72 DPI)

Figure 20a. a) Inverted source wavelet in the frequency band 3-20Hz and b) its spectrum.

2542x1381mm (72 x 72 DPI)

Figure 20b. a) Inverted source wavelet in the frequency band 3-20Hz and b) its spectrum.

2462x1388mm (72 x 72 DPI)

Figure 21a. a) Initial velocity; inverted velocity by b) L2 norm and c) ATI misfit function d) final result after L2 norm inversion refinement of c).

2464x1359mm (72 x 72 DPI)

Figure 21b. a) Initial velocity; inverted velocity by b) L2 norm and c) ATI misfit function d) final result after L2 norm inversion refinement of c).

2464x1359mm (72 x 72 DPI)

Figure 21c. a) Initial velocity; inverted velocity by b) L2 norm and c) ATI misfit function d) final result after L2 norm inversion refinement of c).

2464x1359mm (72 x 72 DPI)

Figure 21d. a) Initial velocity; inverted velocity by b) L2 norm and c) ATI misfit function d) final result after L2 norm inversion refinement of c).

2464x1359mm (72 x 72 DPI)

Figure 22 Comparison of the common shot record for the measured data (red) and modeled data by initial model (black).

1471x1471mm (72 x 72 DPI)

Figure 23 Comparison of the common shot record for the measured data (red) and modeled data by result inverted by L2 norm (black).

1471x1471mm (72 x 72 DPI)

Figure 24 Comparison of the common shot record for the measured data(red) and modeled data by result inverted by ATI plus L2 norm(black).

1471x1471mm (72 x 72 DPI)

Figure 25a. Zoom-in view of the area denoted by blue rectangles in Figure 23 and Figure 24: comparison of the common shot record for the measured data (red) and modeled data by result inverted by a) L2 norm (black) b) ATI plus L2 norm (black).

1471x1471mm (72 x 72 DPI)

Figure 25b. Zoom-in view of the area denoted by blue rectangles in Figure 23 and Figure 24: comparison of the common shot record for the measured data(red) and modeled data by result inverted by a) L2 norm (black) b) ATI plus L2 norm(black).

1471x1471mm (72 x 72 DPI)

Figure 26 Image by initial velocity (Figure 21a)

1411x745mm (72 x 72 DPI)

Figure 27 Image by inverted velocity using L2-norm misfit function(Figure 21b).

1411x745mm (72 x 72 DPI)

Figure 28 Image by inverted velocity using ATI plus L2-norm misfit function(Figure 21d).

1411x745mm (72 x 72 DPI)

Figure 29 Angle domain common image gathers for initial velocity shown(Figure 21a).

1411x735mm (72 x 72 DPI)

Figure 30 Angle domain common image gathers for inverted velocity by L2-norm misfit function(Figure 21b).

1411x735mm (72 x 72 DPI)

Figure 31 Angle domain common image gathers for inverted velocity by ATI plus L2-norm misfit function(Figure 21d).

1411x735mm (72 x 72 DPI)

DATA AND MATERIALS AVAILABILITY

Data associated with this research are confidential and cannot be released.