GEOPHYSICS®

Towards autonomous event identifications in wave-equation traveltimes inversion

Journal:	<i>Geophysics</i>
Manuscript ID	GEO-2020-0940.R2
Manuscript Type:	Technical Paper
Keywords:	
Manuscript Focus Area:	Seismic Inversion, Seismic Velocities and Traveltimes

SCHOLARONE™
Manuscripts

Towards autonomous events identification in wave-equation
traveltime inversion

Fuqiang Chen¹ and Daniel Peter¹

¹*Division of Physical Science and Engineering, King Abdullah University of Science and
Technology, Thuwal, 23955-6900, Saudi Arabia*

(July 21, 2021)

Running head: **Autonomous WT inversion**

ABSTRACT

We present a method to automatically relate events between observed and synthetic data for wave-equation traveltime (WT) inversion. The scheme starts with local similarity measurements by applying cross-correlation to localized traces. We then develop a differentiable alternative of the argmax function. By applying the differentiable argmax to each time slice of the local similarity map, we build a traveltime difference function but keep this process differentiable. WT inversion requires only the traveltime difference of related events through a phase shift. Thus, we must reject events that are not apparently related between observed and synthetic data. The local similarity map implies the possibility of doing so but shows abrupt changes with time and offset. To mitigate this problem, we introduce a dynamic programming algorithm to define a warping function. Wave packets between observed and synthetic data are assumed to be related if they are connected by this warping function, and they also exhibit high local similarity. Only such related events are considered in the subsequent calculation of misfit and adjoint source. Numerical examples demonstrate that the proposed method successfully retrieves an informative model from roughly selected data.

In contrast, WT inversion based on cross-correlation or deconvolution fails to do so.

INTRODUCTION

Full-waveform inversion (FWI) has drawn much attention from academia and industry since its formulation (Tarantola, 1984). Although it is theoretically characterized by many advantages over existing methods in terms of model parameter estimation, setbacks in field data inversions prevent the achievement of its extraordinary potential partially because the high non-linearity between model and data gives rise to numerous local minima in the associated misfit function (Symes, 2008; Symes et al., 2020). This pending problem restricts the application of FWI to many field datasets. A way of alleviating the local minima issue is the construction of well-behaved misfit functions based on optimal transport (Engquist et al., 2016; Yang et al., 2018; Métivier et al., 2016; Chen and Peter, 2018). Herein, we follow wave-equation traveltime (WT) inversion, another direction motivated by the observation of relaxed non-linearity in traveltime inversion. WT inversion replaces the original forward modeling engine in ray-based tomography with the wave equation, thereby combining the robustness of traveltime inversion and the strength of full wave-equation modeling.

Full wave-equation modeling generates multiple events in predicted data. WT inversion requires only those events related through phase shift to focus on traveltime matching. Tailored misfit functions based on cross-correlation (Luo and Schuster, 1991; Van Leeuwen and Mulder, 2010; Wang et al., 2014; Luo et al., 2016) and deconvolution (Luo and Sava, 2011; Warner and Guasch, 2016; Zhu and Fomel, 2016) have been proposed to calculate the traveltime difference. Without appropriate data selection in advance, cross-correlation and stationary deconvolution are insufficient to obtain meaningful traveltime difference measurements for WT inversion. The dynamic time warping (DTW) algorithm is another long-established method employed for calculating the traveltime difference between two

sequences. This non-linear operator allows imposing a constraint on time shifts to maintain continuity (Sakoe and Chiba, 1978; Hale, 2013). It provides robust measurements in case data are noise-contaminated or the traveltimes differences are beyond the constant phase-shift. The DTW algorithm locally estimates the traveltime difference. Therefore, it yields a connective function that quantifies the dissimilarity between observed and synthetic data by the phase shift and other general warping functions. Inversion with a misfit function based on the DTW has also been developed (Ma and Hale, 2013). We categorize this method into WT-type inversions. It is no clear evidence that WT-type inversions are advantageous to convert the general warping functions into model perturbation. To only include related events through a phase shift in misfit calculation, manual or automatic event identification is required.

This paper presents traveltime misfit measurements using an autonomous data selection mechanism, thereby reducing the ambiguity inherent to the estimation of traveltime difference due to subjective picking. Our new approach follows recent studies on building automatic workflows for WT-type inversions (Maggi et al., 2009; Tape et al., 2010), where the authors mostly used statistical methods. In the following, we present a short review of existing WT-type inversion methods. Then, we define a new misfit function based on the differentiable argmax function and local cross-correlation (LCC). Subsequently, we explain how the proposed misfit automatically finds related events. Finally, we demonstrate its potential using both the Chevron blind test and Volve field data.

THEORY

Given synthetic and observed traces $f(t)$ and $g(t)$, the connective function $c(t)$ can be defined as follows:

$$c(t) = (f \star g)(t), \quad (1)$$

or

$$f(t) = (g * c)(t), \quad (2)$$

where $\star$ and $*$ represent cross-correlation and convolution, respectively. If $f(t)$ and $g(t)$ are related through a phase shift, $c(t)$ explicitly carries information regarding the traveltime difference between them, which can be extracted using an activation function $y(t)$. A prototype misfit function ε for WT inversion is then defined as follows:

$$\varepsilon := \frac{\sum_{i=1}^{nl} (y_i c_i)^2}{\sum_{i=1}^{nl} c_i^2}, \quad (3)$$

where nl denotes the number of elements in discrete $c(t)$ (Luo and Schuster, 1991; Van Leeuwen and Mulder, 2010; Luo and Sava, 2011; Luo et al., 2016; Warner and Guasch, 2016; Zhu and Fomel, 2016). The misfit definition in the original WT inversion (Luo and Schuster, 1991) is equivalent to equation 3. Cross-correlation is a widely used method for estimating traveltime difference. Deconvolution is also widely used in seismic data processing for different purposes including the estimation of traveltime differences. Non-stationary deconvolution or adaptive matching filter is another procedure for representing the traveltime differences. The connective function from non-stationary deconvolution is a high-dimensional map (Zhu and Fomel, 2016) and thus formally similar to the connective function defined by LCC herein.

Differentiable argmax function

A simple activation function is an absolute function with its vertex at zero time lag (Luo and Sava, 2011; Warner and Guasch, 2016; Zhu and Fomel, 2016). Instead of explicitly measuring the traveltime difference, it penalizes the connective function such that the zero lag in $c(t)$ approaches the vertex of the absolute function as the misfit is minimized. A natural activation function that directly performs this requirement is the argmax function. It is piece-wise smooth and thus globally non-differentiable. Here, we define a differentiable argmax function as

$$\widehat{\operatorname{argmax}}_i x[i] \equiv \frac{\sum_{i=1}^{nl} i e^{\alpha x[i]}}{\sum_{i=1}^{nl} e^{\alpha x[i]}}, \quad (4)$$

where α is a parameter that controls the compactness of the output. The limit of the differentiable argmax equals the original one as α approaches positive infinity. The derivative of the output of this operator with its arguments is as follows:

$$\frac{\partial \widehat{\operatorname{argmax}}_i x[i]}{\partial x[i]} = \alpha \frac{\left(i - \widehat{\operatorname{argmax}}_i x[i] \right) e^{\alpha x[i]}}{\sum_{i=1}^{nl} e^{\alpha x[i]}}. \quad (5)$$

The calculations in equations 4 and 5 can result in an overflow problem. An unconditionally stable calculation is thus as follows:

$$\widehat{\operatorname{argmax}}_i x[i] = \frac{\sum_{i=1}^{nl} i e^{\alpha(x[i]-x')}}{\sum_{i=1}^{nl} e^{\alpha(x[i]-x')}}}, \quad (6)$$

and

$$\frac{\partial \widehat{\operatorname{argmax}}_i x[i]}{\partial x[i]} = \alpha \frac{\left(i - \widehat{\operatorname{argmax}}_i x[i] \right) e^{\alpha(x[i]-x')}}{\sum_{i=1}^{nl} e^{\alpha(x[i]-x')}}}, \quad (7)$$

where $x' = \max_i x[i]$. Compared to the absolute function, the new activation function reduces the interference from elements not associated with the traveltimes in the connective function.

Revisiting local cross-correlation

Challenges arise when using equation 1 or 2 to estimate the traveltimes difference because $f(t)$ and $g(t)$ are generally not globally related through a phase shift. Without proper data selection, the connective function $c(t)$ is unable to reflect kinematic differences. This breaks the principles of WT inversion. To determine the apparently related wave packets, we define a local connective function as

$$C[i, j] := \sum_{k=1}^{nw} \mathcal{W}[i, k] f[k] \mathcal{W}[i + j, k] g[k], \quad (8)$$

where nw is the length of a window and $\mathcal{W}[i, :]$ is any window function centered at i to localize $f[:]$ and $g[:]$. Here, $f[:]$ refers to all elements in f . Equation 8 is a discrete definition of LCC equivalent to the original one (Hale, 2006). The non-normalized amplitude in windowed $f(t)$ and $g(t)$ can hinder the local similarity measurements. By introducing two auxiliary matrices:

$$\mathcal{W}_f[i, k] = \mathcal{W}[i, k] f[k], \quad (9a)$$

$$\mathcal{W}_g[i, k] = \mathcal{W}[i, k] g[k], \quad (9b)$$

the normalized LCC can be defined as follows:

$$C[i, j] = \frac{\sum_{k=1}^{nw} \mathcal{W}_f[i, k] \mathcal{W}_g[i + j, k]}{\sqrt{\sum_{k=1}^{nw} \mathcal{W}_f^2[i, k] \sum_{k=1}^{nw} \mathcal{W}_g^2[i + j, k] + \delta}}, \quad (10)$$

where δ is a constant to prevent zero division. The derivative of $\mathcal{C}[i, j]$ in equation 10 with $f[k]$ is:

$$\frac{\partial \mathcal{C}[i, j]}{\partial f[k]} = \mathcal{P}^{-1}[i, j] \mathcal{W}[i, k] \left(\mathcal{W}_g[i + j, k] - \mathcal{W}_f[i, k] \mathcal{C}[i, j] \mathcal{Q}[i, j] \right), \quad (11)$$

where

$$\mathcal{P}[i, j] = \sqrt{\sum_{k=1}^{nw} \mathcal{W}_f^2[i, k] \sum_{k=1}^{nw} \mathcal{W}_g^2[i + j, k] + \delta}$$

$$\mathcal{Q}[i, j] = \mathcal{P}^{-1}[i, j] \sum_{k=1}^{nw} \mathcal{W}_g^2[i + j, k]$$

$\mathcal{P}$ and $\mathcal{Q}$ are two auxiliary variables that save values for the calculation of equation 11. Such approaches are generally computationally expensive for real-time processing. However, to successfully mitigate the inverse problem associated with wave-equations, the initial model should not be very far from the plausible model, which means nl is just a fraction of the record length. Computations are always parallelized to accelerate large-scale non-linear inverse problems. Equations 8 and 11 are suitable for acceleration by multi-processing programming. Thus, the calculation of equations 8 and 11 introduces only marginal costs.

Misfit and adjoint source

By introducing

$$u[i] = \widehat{\operatorname{argmax}}_j \mathcal{C}[i, j], \quad (12)$$

we now define a new misfit function as

$$\varepsilon := \sum_{i=1}^{nt} u^2[i], \quad (13)$$

where nt is the number of samples in a trace. Note that the contribution of every sample to ε is independent. This implies that some samples in the calculation of misfit function

can be rejected. Applying the chain rule, we obtain the adjoint source as

$$\frac{\partial \varepsilon}{\partial f[k]} = 2 \sum_{i=1}^{nw} u[i] \sum_{j=1}^{nl} \frac{\partial u[i]}{\partial \mathcal{C}[i, j]} \frac{\partial \mathcal{C}[i, j]}{\partial f[k]}. \quad (14)$$

The simple implementation of equations 14 and 8 requires a triple-nested loop. They are numerically accurate up to the machine epsilon. When calculating the misfit and adjoint source, we can introduce a conditional statement to mute wave packets with low similarity solely based on the LCC map. However, cross-correlation is an adequate but not robust method to measure the time shifts (Hale, 2013). The core of LCC is still a correlation process running over the localized signals. Therefore, it lacks the required robustness to maintain the slow change of time shifts between adjacent samples. By re-writing equation 13 as

$$\varepsilon := \max_{u[i]} \sum_{i=1}^{nt} \mathcal{C}[i, u[i]], \quad (15)$$

we note that $u[i]$ determined by argmax admits the arbitrary change between $u[i]$ and $u[i - 1]$. In the following, we introduce the DTW algorithm to solve equation 15 efficiently under a constraint on $|u[i] - u[i - 1]|$.

Dynamic time warping

The original DTW algorithm can be adapted here with only one modification, i.e., replacing the min operator with max in a sweeping template because the error matrix calculated by LCC prefers the larger values to get tracked. The basic implementation consists of two steps (Sakoe and Chiba, 1978; Hale, 2013): In the first step, we use a template to sweep over the error matrix $\mathcal{C}$ to build another matrix that saves all paths satisfying certain constraints. The second step will sweep reversely by following the same template to find the optimal path. The time complexity for these two steps is $\mathcal{O}(nt \times nl)$ and $\mathcal{O}(nt)$, respectively. Following

the formula proposed by Hale (2013), the first step with a three-point sweeping template reads as follows

$$\mathcal{D}[1, j] = \mathcal{C}[1, j]$$

$$\mathcal{D}[i, j] = \mathcal{C}[i, j] + \max \begin{cases} \mathcal{D}[i-1, j-1] \\ \mathcal{D}[i-1, j] \\ \mathcal{D}[i-1, j+1] \end{cases}, \quad (16)$$

for $i=2, 3, \dots, nt$.

where $\mathcal{D}$ denotes the array of distances formed by forward sweeping. The second step, termed backtracking, is a process of searching over the accumulated matrix $\mathcal{D}$ and can be formulated as follows:

$$u[nt] = \underset{j}{\operatorname{argmax}} \mathcal{D}[nt, j]$$

$$u[i-1] = \underset{j \in \{u[i] \pm 1, u[i]\}}{\operatorname{argmax}} \mathcal{D}[i-1, j], \quad (17)$$

for $i = nt, \dots, 3, 2$.

where array u denotes the effective time shifts. Warping shifts resulting from the three-point template satisfy $|u[i] - u[i-1]| \leq 1$ (Hale, 2013). The robustness of DTW in travel-time difference estimation can be further improved by considering the traveltime difference variation in the offset dimension. An approximate but very efficient algorithm is entitled tree-sequential dynamic warping (Hale, 2013). In brief, this algorithm alternately applies sweeping templates in the time-lag and offset-lag planes. The DTW algorithm enhanced by multi-dimensional sweeping is especially useful for selecting the related reflection events as discussed below or when data are contaminated by noise (Hale, 2013).

NUMERICAL EXAMPLES

In this section, we propose a two-parameter inverse problem to investigate the landscape of misfit functions focusing on traveltimes differences. Then, we use Chevron blind test data to demonstrate that proper data preparation is essential for the deconvolution method to evaluate the traveltimes difference. Finally, we illustrate the capability of the proposed method in automatic data selection for WT inversion.

A toy example

The single-parameter inverse problem is commonly used to demonstrate the power of a misfit function in annihilating the local minima. It can be extended to a two-parameter inverse problem if the governing equation becomes $f(t) = r(t; t_0, f_m) + r(t; t_1, f_m)$, where $r(t; t_0, f_m)$ denotes the Ricker wavelet with the peak frequency f_m and peak amplitude at time t_0 . Figure 1 displays observed and two instances of synthetic data, with a peak frequency $f_m = 10.0$ Hz. Figure 2a–2c shows the misfit measured by $\mathcal{L}^2$ -norm, global cross-correlation, and deconvolution, respectively. Figure 2d–2f shows the result from LCC, where the absolute function (Figure 2d), the differentiable argmax with $\alpha = 1$ (Figure 2e) and $\alpha = 10$ (Figure 2f) are used as activation functions. All misfit functions are normalized in $[0, 1]$. Figure 2a–2c shows that deconvolution yields improved misfit measurements compared to cross-correlation and $\mathcal{L}^2$ -norm even all methods present severe local minima issues. One interesting observation in Figure 2b and 2c is that the misfit functions are still convex along the direction indicated by the solid arrows where two events in synthetic data are translated by the same perturbations. Thus, the two-parameter model collapses to the single-parameter wherein cross-correlation and deconvolution are appropriate for traveltimes

difference measurements. We theoretically explain why deconvolution fails to provide convex measurements for all predicted models in this two-spike example in Appendix A. Figure 2d–2f shows that misfit functions based on local similarities successfully mitigate the local minima because only traveltimes differences from related events are counted. The length of the Gaussian window and the range of time shift for LCC are 0.3 s and $|\tau| \leq 0.35$ s, respectively. Moreover, Figure 2e–2f shows that the differentiable argmax activation function with a larger α works better when the predicted model is far from the true model. Because the local similarity map is always normalized in $[-1, 1]$, we consider that the value of α in this testing has generality. We take $\alpha = 10$ for the differentiable argmax in Chevron blind test and Volve field-data examples. This operator is similar to the max function used in equation 2 in the original WT inversion method (Luo and Schuster, 1991). The proposed argmax function is differentiable; therefore it is natural to apply differentiation rules for the derivation of the adjoint source. Another two hyper-parameters to calculate local travel-time differences are the length of the window and the range of time shift. We roughly take the longest wavelength as the window length. The maximum magnitude of the traveltimes differences determines the range of time shifts.

Chevron blind test data

The Chevron 2014 datasets are widely used to test different FWI algorithms (Qin et al., 2015; Vigh et al., 2016; Xu and Macesanu, 2016; Warner and Guasch, 2016; Métivier et al., 2016), where the observed data are generated by solving the elastic wave equation with free surface boundary conditions. An accurate wavelet and a good initial model shown in Figure 3 are available. We mute the noise in observed data before the first arrival. This practical pre-processing is helpful but not critically important for successful WT inversion.

We first demonstrate how the proposed strategy automatically selects related events. Figure 4 shows three observed traces with different offsets and their synthetic counterparts from the initial model. The corresponding LCC maps are shown in Figure 5. This local map enables the determination of a traveltime difference function using the argmax operator along the time axis. It is unreasonable to involve all samplings in the misfit calculation. Instead of manually selecting related events through a phase shift, we propose two automatic strategies based on the LCC map. The first, constant thresholding, rejects samples with lower local similarity than a threshold β . The second, joint thresholding, further rejects samples not connected by the warping path. The red and blue dots in Figure 5 denote traveltime differences determined by the constant thresholding of $\beta = 0.9$ and the DTW algorithm, respectively. This demonstrates that the constant thresholding strategy only based on the LCC map generates abrupt changes. In contrast, changes between adjacent blue dots are strictly bounded. Graphically, the joint thresholding recognizes samplings connected by both red and blue dots as related wave packets.

To demonstrate the importance of events identification for a successful WT-type inversion, we compare two methods based on different connective functions: deconvolution and the proposed one. Because WT-type inversions are still challenging for reflections, we only use early arrivals in this comparison. Therefore, observed and synthetic data out of the shade of green are muted as shown in Figure 6. We consider this pre-processing as a rough data selection. Figure 7 illustrates the adjoint sources by deconvolution, LCC without data selection, LCC with constant thresholding of $\beta = 0.5$, LCC with constant thresholding of $\beta = 0.9$, and joint thresholding with $\beta = 0.9$. The corresponding gradients are shown in Figure 8, where the frequent changes of polarity in Figure 8a–8d indicate incorrect traveltime difference measurements. The deconvolution problem is iteratively solved by a modified

conjugate gradient method, in which we use Tikhonov regularization to ensure the connective function unique and smooth. Two parameters need to be decided: the length of the connective function nl and the parameter controlling the regularization strength λ . We take $nl = 0.6$ s based on the maximum phase shift between observed and synthetic traces. Finding an optimal value for λ in this application is not trivial. We performed a series of experiments by scaling λ a factor of 4 until the gradient stops showing a significant difference. None of λ can give correct traveltimes difference measurements with the roughly selected data. The λ for Figure 8b is 16 times larger than the one for Figure 8a. This experiment demonstrates that appropriate data selection is necessary for WT inversion based on the deconvolution misfit.

We apply the proposed WT inversion with the joint thresholding to two stages: 2.0 Hz–3.0 Hz and 2.0 Hz–4.0 Hz. Figure 9 shows the accumulated updates from these two successive stages and indicates that the updated perturbation is bounded by 10%. Therefore, the differences between migration results from initial and inverted models are not significant as shown in Figure 10. However, we can still observe the improvement from the model update. For example, the artifacts in Figure 10a indicated by the orange arrow are attenuated in Figure 10b due to increasingly accurate velocity. Furthermore, the continuity of horizons indicated by red arrows is improved in Figure 10. Figure 11 illustrates the comparison in the data domain where observed and synthetic datasets are interleaved to compare the phase alignment by offset groups. Stripes with stronger reflections indicate observed data. The synthetic data in Figure 11a and 11b are generated based on initial and updated models, respectively. Figure 11a shows apparent phase misalignment as offset greater than 3.0 km. Figure 11b, shows a significant reduction of phase mismatching, indicating that the proposed WT inversion successfully recovers the residual long-wavelength component in the case of rough

data selection.

Finally, we continue the waveform inversion with the misfit function based on $\mathcal{L}^2$ -norm to fit the reflections. This takes three stages: 4.0 Hz–6.0 Hz, 5.0 Hz–8.0 Hz, and 7.0 Hz–10.0 Hz. Figure 12 shows the final inverted result. Figure 13 illustrates the well-log comparison wherein the predicted velocity profiles approach the true one. A multi-scale workflow with FWI with the misfit function based on $\mathcal{L}^2$ -norm yields less plausible results when starting from the initial model in Figure 3. This has already been reported in a previous study (Xu and Macesanu, 2016).

The North Sea Volve field data

The purpose of this example is to test the ability of the joint thresholding strategy to identify related events for WT inversion from field data. Figure 14 illustrates the surface layout of the sources (blue and orange), receivers (red), and induced CDP line (green). We follow the CDP line to extract the reference model (Figure 15) from the velocity cube for depth migration. Figure 16a shows a receiver gather, where the synthetic data are generated based on the reference model and a band-limited wavelet in 3.0 Hz–5.0 Hz, then the synthetic and observed data are also re-arranged in an interleaved style to demonstrate the residual traveltime mismatch highlighted in the red quadrilaterals. In this experiment, we use 240 shots (orange dots in Figure 14) for the proposed WT inversion. The initial model is a smoothed version of the reference model. We estimate the wavelet every 30 shots starting from the far-field signature (Sun et al., 2014). Besides, we apply a simple amplitude correction scheme (Crase et al., 1990) to ensure that the wavefield propagation for the observed data complied with cylindrical spreading. The inversion is finalized by two

stages: 2.0 Hz–4.0 Hz and 3.0 Hz–5.0 Hz. Figure 16b, where the synthetic data are generated based on the inverted model, clearly shows that the phase alignment inside the red boxes is improved by comparing with Figure 16a. Figure 17 shows the estimated perturbation from this inversion.

The essence of seismic exploration is decoding the information carried by reflections. In the following, we explore the possibility of extending the proposed strategy to identify related events through a phase shift from reflections. Figure 18a and 18b shows observed and synthetic receiver gathers with the frequency band in 2.0 Hz–3.0 Hz. Noise is visibly present in the observed data. Contrary to early arrivals, the correlation of reflection arrivals from observed and synthetic data is generally low. It makes the traveltime difference estimated by the trace-by-trace DTW changes abruptly with offset. Therefore we use the tree-sequential DTW algorithm to increase the smoothness along the offset axis. Figure 19a shows the estimated time shifts from LCC maps. Figure 19b illustrates the result from joint thresholding of $\beta = 0.3$, where the minimum value of time-shift indicates the samples rejected by following the joint thresholding, and it demonstrates the proposed strategy can still identify the related early arrivals from data with such noise level and fails to find the related reflections. This experiment was repeated with datasets in 2.0 Hz–4.0 Hz as shown in Figure 20. Figure 21 demonstrates that the performance of the proposed strategy improves for reflections with a higher signal-noise ratio. The thresholding parameter significantly impacts the adjoint source, thus the inversion result as well. Compared to the threshold value used in the Chevron blind test, we select much smaller values for this example due to the lower correlation in the LCC map. We do not further verify the possibility of identifying the related reflections automatically with higher frequencies. Beyond the low frequencies, the kernel of WT inversion also should be changed such that the traveltime mismatch will

be backward propagated along the reflection wave-path in gradient calculation (Xu et al., 2012).

CONCLUSION

We demonstrated that data selection is required for successful WT-type inversions based on cross-correlation and deconvolution to accurately estimate the traveltime difference. We then presented a new misfit function combining the normalized local cross-correlation and dynamic time warping. This new misfit function automatically relates events from roughly selected data and considers only the traveltime difference between related events through a phase shift. It is an improvement compared to other WT-type inversions that need manual identification of related events before the inversion. Synthetic and field data examples demonstrate that the proposed method can retrieve residual long-wavelength components from an unfavorable data selection. This is the case where we show that WT inversion based on deconvolution yields implausible estimations. Our proposed strategy also shows the potential for automatically identifying related reflections.

ACKNOWLEDGMENT

We acknowledge the Supercomputing Laboratory at King Abdullah University of Science & Technology (KAUST) for providing resources that contributed to the research results reported within this manuscript. We thank two anonymous reviewers for their insightful comments to improve our paper. We also thank Jeffrey Shragge, Colin Farquharson, and the anonymous associate editor for their helpful feedback. We would like to thank Equinor and the Volve license partners for making the Volve field dataset available for scientific

research. We would like to thank Chevron for making the SEG 2014 Chevron FWI synthetic available.

APPENDIX A

Stationary deconvolution of two-spike signals

Suppose that the observed data are as follows:

$$g(t) = r(t - t_0) + r(t - t_1), \quad (\text{A-1})$$

where $r(t)$ is any function that can be represented by Fourier series and $\mathbf{t} = (t_0, t_1)$ is the true model parameter. The synthetic data are

$$f(t) = r(t - \tau_0) + r(t - \tau_1). \quad (\text{A-2})$$

Applying the time shifting property of Fourier transform, we have

$$\begin{aligned} \hat{g}(\omega) &= \hat{r}(\omega)(e^{-j\omega t_0} + e^{-j\omega t_1}), \\ \hat{f}(\omega) &= \hat{r}(\omega)(e^{-j\omega \tau_0} + e^{-j\omega \tau_1}), \end{aligned} \quad (\text{A-3})$$

where $\hat{f}(\omega)$ is the Fourier transform of $f(t)$. To explore the property of deconvolution-based misfit in the two-spike problem, we define the following connective function:

$$\hat{c}(\omega) := \frac{\hat{f}}{\hat{g}} = \frac{e^{-j\omega \tau_0} + e^{-j\omega \tau_1}}{e^{-j\omega t_0} + e^{-j\omega t_1}}. \quad (\text{A-4})$$

Note that the connective function $\hat{c}(\omega)$ also builds a linear time-invariant (LTI) system by defining the convolution process. Let $z(\tau) = e^{-j\omega \tau}$ and substitute it into equation A-4; we obtain

$$\hat{c}(\omega) = \frac{z(\tau_0) + z(\tau_1)}{z(t_0) + z(t_1)}. \quad (\text{A-5})$$

In this appendix, we only discuss two special cases. For the first one indicated by the solid arrow in Figure 2c, we have $\tau_0 = t_0 + \Delta$, $\tau_1 = t_1 + \Delta$. substituting them into equation A-5, we obtain

$$\hat{c}(\omega) = z(\Delta) = e^{-j\omega\Delta}. \quad (\text{A-6})$$

The output from this LTI system preserves the amplitude spectrum. The phase spectrum is $-\omega\Delta$ and the phase delay is $-\Delta$, which is frequency-independent. Therefore, we can obtain a convex measurement of traveltime difference between input and output from such a system in the time domain.

For the second case indicated by the dashed arrow in Figure 2c, we have $\tau_0 = t_0 + \Delta$, $\tau_1 = t_1 - \Delta$. Substituting them into equation A-5, we obtain

$$\hat{c}(\omega) = \frac{z(t_0)z(\Delta) + z(t_1)z(-\Delta)}{z(t_0) + z(t_1)}. \quad (\text{A-7})$$

Using the trigonometric identities, we simplify equation A-7 as:

$$\begin{aligned} \hat{c}(\omega) &= \frac{\cos(\omega\Delta) + \cos(\omega(t_1 - t_0 - \Delta))}{1 + \cos(\omega(t_1 - t_0))} \\ &= \cos(\omega\Delta) + \sin(\omega\Delta) \frac{\sin(\omega(t_1 - t_0))}{1 + \cos(\omega(t_1 - t_0))} \\ &= \cos(\omega\Delta) + \sin(\omega\Delta) \tan\left(\frac{\omega}{2}(t_1 - t_0)\right). \end{aligned} \quad (\text{A-8})$$

Because the phase delay is frequency-dependent in this case, it is unfeasible for stationary deconvolution to obtain a convex measurement of traveltime difference between input and output from such a system in the time domain.

REFERENCES

- Chen, F., and D. Peter, 2018, Constructing misfit function for full waveform inversion based on sliced Wasserstein distance: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts, We A12 03.
- Cruse, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: *GEOPHYSICS*, **55**, 527–538.
- Engquist, B., B. D. Froese, and Y. Yang, 2016, Optimal transport for seismic full waveform inversion: arXiv preprint arXiv:1602.01540v2.
- Hale, D., 2006, An efficient method for computing local cross-correlations of multi-dimensional signals: Colorado School of Mines Center for Wave Phenomena (CWP) Report 544.
- , 2013, Dynamic warping of seismic images: *GEOPHYSICS*, **78**, no. 2, S105–S115.
- Luo, S., and P. Sava, 2011, *in* A deconvolution-based objective function for wave-equation inversion: 2788–2792.
- Luo, Y., Y. Ma, Y. Wu, H. Liu, and L. Cao, 2016, Full-traveltime inversion: *GEOPHYSICS*, **81**, no. 5, R261–R274.
- Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: *GEOPHYSICS*, **56**, 645–653.
- Ma, Y., and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion: *GEOPHYSICS*, **78**, no. 6, R223–R233.
- Maggi, A., C. Tape, M. Chen, D. Chao, and J. Tromp, 2009, An automated time-window selection algorithm for seismic tomography: *Geophysical Journal International*, **178**, 257–281.
- Métivier, L., R. Brossier, Q. Mérigot, 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**, 345–377.

Qin, B., T. Allemand, and G. Lambaré, 2015, Full waveform inversion using preserved amplitude reverse time migration: 85th Annual International Meeting, SEG, Expanded Abstracts, 1252–1257.

Sakoe, H., and S. Chiba, 1978, Dynamic programming algorithm optimization for spoken word recognition: *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **26**, 43–49.

Sun, D., K. Jiao, X. Cheng, D. Vigh, and R. Coates, 2014, Source wavelet estimation in full waveform inversion: 84th Annual International Meeting, SEG, Expanded Abstracts, 1184–1188.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical Prospecting*, **56**, 765–790.

Symes, W. W., H. Chen, and S. E. Minkoff, 2020, Full-waveform inversion by source extension: Why it works: 90th Annual International Meeting, SEG, Expanded Abstracts, 765–769.

Tape, C., Q. Liu, A. Maggi, and J. Tromp, 2010, Seismic tomography of the southern California crust based on spectral-element and adjoint methods: *Geophysical Journal International*, **180**, 433–462.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *GEOPHYSICS*, **49**, 1259–1266.

Van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation traveltimes tomography: *Geophysical Journal International*, **182**, 1383–1394.

Vigh, D., K. Jiao, X. Cheng, D. Sun, and W. Lewis, 2016, Earth-model building from

- shallow to deep with full-waveform inversion: The Leading Edge, **35**, 1025–1030.
- Wang, H., S. C. Singh, and H. Calandra, 2014, Integrated inversion using combined wave-equation tomography and full waveform inversion: Geophysical Journal International, **198**, 430–446.
- Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: GEOPHYSICS, **81**, no. 6, R429–R445.
- Xu, K., and C. Macesanu, 2016, First-arrival waveform inversion using low-frequency regenerated data: 86th Annual International Meeting, SEG, Expanded Abstracts, 5665–5669.
- Xu, S., D. Wang, F. Chen, Y. Zhang, and G. Lambare, 2012, Full waveform inversion for reflected seismic data: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts, W024.
- 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.
- Zhu, H., and S. Fomel, 2016, Building good starting models for full-waveform inversion using adaptive matching filtering misfit: GEOPHYSICS, **81**, no. 5, U61–U72.

LIST OF FIGURES

- 1 The observed (black) with $\mathbf{t} = (0.5, 1.1)$ and two instances of synthetic traces with
- 2 $\mathbf{t} = (0.65, 0.95)$ and $\mathbf{t} = (0.35, 1.25)$, respectively.
- 3 Misfit functions based on $\mathcal{L}^2$ -norm (a), cross-correlation (b), deconvolution (c) and
- 4 normalized local cross-correlation (d-f). The absolute function is used for (b-d) and differ-
- 5 entiable argmax with $\alpha = 1.0$ and $\alpha = 10.0$ are used for (e-f) as the activation function.
- 6 Initial starting model released with the Chevron blind test datasets.
- 7 The observed and synthetic traces to exemplify automatic event identifications.
- 8 The corresponding LCC maps associated with the observed and synthetic traces
- 9 in Figure 4. The red and blue dots are determined by the constant thresholding of $\beta = 0.9$
- 10 and DTW, respectively. The overlapped parts indicate the selected wave packets by the
- 11 joint thresholding strategy.
- 12 An example of rough data selection for WT-type inversion.
- 13 The adjoint sources from different traveltime difference measurements: (a-b) decon-
- 14 volution with different regularization strength, (c) LCC with no data selection, (d) constant
- 15 thresholding of $\beta = 0.5$, (e) constant thresholding of $\beta = 0.9$ and (f) joint thresholding with
- 16 $\beta = 0.9$.
- 17 The results of gradient from different adjoint sources as shown in Figure 7. All
- 18 gradients are normalized in $[-1, 1]$.
- 19 The updated velocity by the proposed WT inversion from roughly selected data.
- 20 Migration results (a) and (b) from the initial model and updated one, respectively.
- 21 The comparison of phase alignment between observed and synthetic data in 2.0 Hz-
- 22 4.0 Hz, where they are re-organized in an interleaved style to compare the phase-matching
- 23 by offset groups.

12 The final inverted model by FWI starting from the result of WT inversion. The
blue segment illustrates where the true velocity is known.

13 The comparison of velocity profiles.

14 A diagram for the 3D survey in the North Sea where only a part of shots (blue)
and one receiver line (red) is illustrated. The 240 shots (orange) are used for inversion.

15 The reference model from the velocity cube obtained by an industry-strength
method.

16 The phase-matching comparison between observed and synthetic data where they
are displayed in an interleaved style. The synthetic data in panels (a) and (b) are generated
based on the reference model and updated one, respectively.

17 The estimated perturbation by WT inversion.

18 The observed (a) and synthetic data (b) with the frequency band in 2.0 Hz–3.0 Hz.

19 (a) Time shifts estimated by LCC+DTW from datasets shown in Figure 18 and
(b) sorted results from joint thresholding with $\beta = 0.3$.

20 The observed (a) and synthetic data (b) with the frequency band in 2.0 Hz–4.0 Hz.

21 (a) Time shifts estimated by LCC+DTW from datasets shown in Figure 20 and
(b) sorted results from joint thresholding with $\beta = 0.5$.

Figure 1: The observed (black) with $\mathbf{t} = (0.5, 1.1)$ and two instances of synthetic traces with $\mathbf{t} = (0.65, 0.95)$ and $\mathbf{t} = (0.35, 1.25)$, respectively.

130x45mm (300 x 300 DPI)

Figure 2a: Misfit functions based on L^2 -norm (a), cross-correlation (b), deconvolution (c) and normalized local cross-correlation (d-f). The absolute function is used for (b-d) and differentiable argmax with $\alpha = 1.0$ and $\alpha = 10.0$ are used for (e-f) as the activation function.

43x38mm (300 x 300 DPI)

Figure 2b: Misfit functions based on L^2 -norm (a), cross-correlation (b), deconvolution (c) and normalized local cross-correlation (d-f). The absolute function is used for (b-d) and differentiable argmax with $\alpha = 1.0$ and $\alpha = 10.0$ are used for (e-f) as the activation function.

43x38mm (300 x 300 DPI)

Figure 2c: Misfit functions based on L^2 -norm (a), cross-correlation (b), deconvolution (c) and normalized local cross-correlation (d-f). The absolute function is used for (b-d) and differentiable argmax with $\alpha = 1.0$ and $\alpha = 10.0$ are used for (e-f) as the activation function.

43x38mm (300 x 300 DPI)

Figure 2d: Misfit functions based on L^2 -norm (a), cross-correlation (b), deconvolution (c) and normalized local cross-correlation (d-f). The absolute function is used for (b-d) and differentiable argmax with $\alpha = 1.0$ and $\alpha = 10.0$ are used for (e-f) as the activation function.

43x38mm (300 x 300 DPI)

Figure 2e: Misfit functions based on L^2 -norm (a), cross-correlation (b), deconvolution (c) and normalized local cross-correlation (d-f). The absolute function is used for (b-d) and differentiable argmax with $\alpha = 1.0$ and $\alpha = 10.0$ are used for (e-f) as the activation function.

43x38mm (300 x 300 DPI)

Figure 2f: Misfit functions based on L^2 -norm (a), cross-correlation (b), deconvolution (c) and normalized local cross-correlation (d-f). The absolute function is used for (b-d) and differentiable argmax with $\alpha = 1.0$ and $\alpha = 10.0$ are used for (e-f) as the activation function.

43x38mm (300 x 300 DPI)

Figure 3: Initial starting model released with the Chevron blind test datasets.

91x37mm (300 x 300 DPI)

Figure 4: The observed and synthetic traces to exemplify automatic event identifications.

130x45mm (300 x 300 DPI)

Figure 5a: The corresponding LCC maps associated with the observed and synthetic traces in Figure 4. The red and blue dots are determined by the constant thresholding of $\beta = 0.9$ and DTW, respectively. The overlapped parts 29 indicate the selected wave packets by the joint thresholding strategy.

58x91mm (300 x 300 DPI)

Figure 5b: The corresponding LCC maps associated with the observed and synthetic traces in Figure 4. The red and blue dots are determined by the constant thresholding of $\beta = 0.9$ and DTW, respectively. The overlapped parts 29 indicate the selected wave packets by the joint thresholding strategy.

58x91mm (300 x 300 DPI)

Figure 5c: The corresponding LCC maps associated with the observed and synthetic traces in Figure 4. The red and blue dots are determined by the constant thresholding of $\beta = 0.9$ and DTW, respectively. The overlapped parts indicate the selected wave packets by the joint thresholding strategy.

58x91mm (300 x 300 DPI)

Figure 6: An example of rough data selection for WT-type inversion.

81x54mm (300 x 300 DPI)

Figure 7a: The adjoint sources from different traveltime difference measurements: (a-b) deconvolution with different regularization strength, (c) LCC with no data selection, (d) constant thresholding of $\beta = 0.5$, (e) constant thresholding of $\beta = 0.9$ and (f) joint thresholding with $\beta = 0.9$.

94x72mm (300 x 300 DPI)

Figure 7b: The adjoint sources from different traveltime difference measurements: (a-b) deconvolution with different regularization strength, (c) LCC with no data selection, (d) constant thresholding of $\beta = 0.5$, (e) constant thresholding of $\beta = 0.9$ and (f) joint thresholding with $\beta = 0.9$.

94x72mm (300 x 300 DPI)

Figure 7c: The adjoint sources from different traveltimes difference measurements: (a-b) deconvolution with different regularization strength, (c) LCC with no data selection, (d) constant thresholding of $\beta = 0.5$, (e) constant thresholding of $\beta = 0.9$ and (f) joint thresholding with $\beta = 0.9$.

94x72mm (300 x 300 DPI)

Figure 7d: The adjoint sources from different traveltime difference measurements: (a-b) deconvolution with different regularization strength, (c) LCC with no data selection, (d) constant thresholding of $\beta = 0.5$, (e) constant thresholding of $\beta = 0.9$ and (f) joint thresholding with $\beta = 0.9$.

94x72mm (300 x 300 DPI)

Figure 7e: The adjoint sources from different traveltime difference measurements: (a-b) deconvolution with different regularization strength, (c) LCC with no data selection, (d) constant thresholding of $\beta = 0.5$, (e) constant thresholding of $\beta = 0.9$ and (f) joint thresholding with $\beta = 0.9$.

94x72mm (300 x 300 DPI)

Figure 7f: The adjoint sources from different traveltim difference measurements: (a-b) deconvolution with different regularization strength, (c) LCC with no data selection, (d) constant thresholding of $\beta = 0.5$, (e) constant thresholding of $\beta = 0.9$ and (f) joint thresholding with $\beta = 0.9$.

94x72mm (300 x 300 DPI)

Figure 8a: The results of gradient from different adjoint sources as shown in Figure 7. All gradients are normalized in $[-1, 1]$.

62x28mm (300 x 300 DPI)

Figure 8b: The results of gradient from different adjoint sources as shown in Figure 7. All gradients are normalized in $[-1, 1]$.

62x28mm (300 x 300 DPI)

Figure 8c: The results of gradient from different adjoint sources as shown in Figure 7. All gradients are normalized in $[-1, 1]$.

62x27mm (300 x 300 DPI)

Figure 8d: The results of gradient from different adjoint sources as shown in Figure 7. All gradients are normalized in $[-1, 1]$.

62x27mm (300 x 300 DPI)

Figure 8e: The results of gradient from different adjoint sources as shown in Figure 7. All gradients are normalized in $[-1, 1]$.

62x27mm (300 x 300 DPI)

Figure 8f: The results of gradient from different adjoint sources as shown in Figure 7. All gradients are normalized in $[-1, 1]$.

62x29mm (300 x 300 DPI)

Figure 9: The updated velocity by the proposed WT inversion from roughly selected data.

91x36mm (300 x 300 DPI)

Figure 10a: Migration results (a) and (b) from the initial model and updated one, respectively.

84x44mm (300 x 300 DPI)

Figure 10b: Migration results (a) and (b) from the initial model and updated one, respectively.

84x44mm (300 x 300 DPI)

Figure 11a: The comparison of phase alignment between observed and synthetic data in 2.0 Hz-4.0 Hz, where they are re-organized in an interleaved style to compare the phase-matching by offset groups.

94x72mm (300 x 300 DPI)

Figure 11b: The comparison of phase alignment between observed and synthetic data in 2.0 Hz-4.0 Hz, where they are re-organized in an interleaved style to compare the phase-matching by offset groups.

94x72mm (300 x 300 DPI)

Figure 12: The final inverted model by FWI starting from the result of WT inversion. The blue segment illustrates where the true velocity is known.

91x37mm (300 x 300 DPI)

Figure 13: The comparison of velocity profiles.

130x45mm (300 x 300 DPI)

Figure 14: A diagram for the 3D survey in the North Sea where only a part of shots (blue) and one receiver line (red) is illustrated. The 240 shots (orange) are used for inversion.

131x58mm (300 x 300 DPI)

Figure 15: The reference model from the velocity cube obtained by an industry-strength method.

133x57mm (300 x 300 DPI)

Figure 16a: The phase-matching comparison between observed and synthetic data where they are displayed in an interleaved style. The synthetic data in panels (a) and (b) are generated based on the reference model and updated one, respectively.

90x60mm (300 x 300 DPI)

Figure 16b: The phase-matching comparison between observed and synthetic data where they are displayed in an interleaved style. The synthetic data in panels (a) and (b) are generated based on the reference model and updated one, respectively.

90x60mm (300 x 300 DPI)

Figure 17: The estimated perturbation by WT inversion.
133x59mm (300 x 300 DPI)

Figure 18a: The observed (a) and synthetic data (b) with the frequency band in 2.0 Hz–3.0 Hz.

90x60mm (300 x 300 DPI)

Figure 18b: The observed (a) and synthetic data (b) with the frequency band in 2.0 Hz–3.0 Hz.

90x60mm (300 x 300 DPI)

Figure 19a: (a) Time shifts estimated by LCC+DTW from datasets shown in Figure 18 and (b) sorted results from joint thresholding with $\beta = 0.3$.

105x63mm (300 x 300 DPI)

Figure 19b: (a) Time shifts estimated by LCC+DTW from datasets shown in Figure 18 and (b) sorted results from joint thresholding with $\beta = 0.3$.

105x63mm (300 x 300 DPI)

Figure 20a: The observed (a) and synthetic data (b) with the frequency band in 2.0 Hz–4.0 Hz.

90x60mm (300 x 300 DPI)

Figure 20b: The observed (a) and synthetic data (b) with the frequency band in 2.0 Hz–4.0 Hz.

90x60mm (300 x 300 DPI)

Figure 21a: (a) Time shifts estimated by LCC+DTW from datasets shown in Figure 20 and (b) sorted results from joint thresholding with $\beta = 0.5$.

105x63mm (300 x 300 DPI)

Figure 21b: (a) Time shifts estimated by LCC+DTW from datasets shown in Figure 20 and (b) sorted results from joint thresholding with $\beta = 0.5$.

105x63mm (300 x 300 DPI)

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.