Local-crosscorrelation elastic full-waveform inversion

Zhen-dong Zhang[1] and Tariq Alkhalifah[1]

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

Email: zhendong.zhang@kaust.edu.sa

(August 25, 2019)

Running head: LCeFWI
ABSTRACT

Full waveform inversion in its classical form is a method based on minimizing the $L_2$ norm of the difference between observed and simulated seismic waveforms at the receiver locations. The objective is to find a subsurface model that reproduces the full waveform including the travel times and amplitudes of the observed seismic data. However, the widely used $L_2$ norm based full waveform inversion faces many issues in practice. The point-wise comparison of waveforms fails when the phase difference between the compared waveforms of the predicted and observed data is larger than a half-cycle. Additionally, amplitude-matching is impractical considering the simplified physics we often use to describe the medium. To avoid these known problems, we develop a novel elastic full waveform inversion algorithm using the local-similarity attribute. It compares two traces within a predefined local time-extension, and thus, is not limited by the half-cycle criterion. The algorithm strives to maximize the local similarities of the predicted and observed data by stretching/squeezing the observed data. Phases instead of amplitudes of the seismic data are used in the comparison. The algorithm compares two data sets locally, and thus, performs better than the global correlation in matching multiple arrivals. Instead of picking/calculating one stationary stretching/squeezing curve, we use a weighted integral to find all possible stationary curves. We also introduce a polynomial-type weighting function, which is determined only by the predefined maximum stretching/squeezing and is guaranteed to be smoothly varying within the extension range. Compared to the previously used Gaussian or Linear weighting functions, the proposed polynomial one has fewer parameters to play around with. A modified synthetic elastic Marmousi model and the North Sea field data are used to verify the effectiveness of the developed approach and also reveal some of its limitations.
Keywords: Elastic FWI, Local similarity, Multicomponent, Local Minima.
INTRODUCTION

Full waveform inversion (FWI) is instrumental in building a high-resolution model for the subsurface (Virieux and Operto, 2009). Conventional FWI attempts to match the waveforms of the predicted data and those observed in the field. It, however, has challenges in its application to field data. The difficulties can be categorized into two components, the insufficient approximations of seismic wave propagation in the real Earth and the often inaccurate initial models used for FWI. By incorporating more physics (such as elasticity, anisotropy and attenuation) in wave extrapolation, the simulated wavefield has a better chance to represent the one in the real Earth. However, including such features in the modeling implies we need to invert for the parameters that describe these features and that introduces all kinds of challenges in resolving parameter tradeoff and nonuniqueness. Finding practical FWI strategies, especially for the initial model building, has drawn a lot of attention from the exploration geophysics community (Hu et al., 2018; Liu et al., 2018; Zhang and Alkhalifah, 2019a). There are several strategies to build such a kinematically-accurate initial model: First, match the kinematic information in the data or image domain such as moveout analysis (Alkhalifah, 1997; Biondi and Almomin, 2014; Sun and Alkhalifah, 2017) and migration velocity analysis (MVA) using differential semblance optimization (DSO) (Symes and Carazzone, 1991; Shen et al., 2005; Wang et al., 2019), reflection based waveform inversion (RWI) (Xu et al., 2012; Wu and Alkhalifah, 2015; Guo et al., 2017; Li et al., 2019; Yao et al., 2019) or full travel-time inversion (FTI) (Luo et al., 2016; Yi et al., 2019). Second, compare arrivals with extensions. The $L_2$ norm objective function requires that the maximum mismatch of the predicted and observed data should not exceed a half-cycle for all arrivals; otherwise, the adjoint source is cycle-skipped. The correlation-based objective function (Van Leeuwen and Mulder, 2008, 2010; Routh et al., 2011; Choi
and Alkhalifah, 2012; Chi et al., 2015; Zhang et al., 2018a; Wu et al., 2019) or adaptive full wave inversion (AWI) (Warner and Guasch, 2014) can compare the arrivals globally within a time window and in some cases are free of cycle-skipping. Third, find a simplified representation of the complex data. The philosophy is to find equivalent data, which preserve the key features of the original ones, but have fewer oscillations in time. The optimal transport approach (Engquist et al., 2016; M´etivier et al., 2016; Yang et al., 2018), the envelope representation (Chi et al., 2014; Wu et al., 2014), the Fourier domain method (Solano et al., 2013) and the skeletonized dispersion curve inversion (Zhang et al., 2016; Zhang and Alkhalifah, 2019b) fall into this category. Fourth, joint inversion. Utilizing different physical models as regularization can reduce the nonlinearity of the inverse problem. Early-arrival traveltime and waveform joint inversion are widely used in practical applications of FWI (Zhang and Chen, 2014; Zhou et al., 2015); Besides, incorporating gravity, electromagnetic and well information into regular FWI is drawing more and more attention (Hu et al., 2009; Zhang et al., 2018b; Zhang and Alkhalifah, 2019a).

In this paper, we introduce a new objective function that measures the difference between the observed and predicted data using local-similarity attributes. The Earth is attenuative, and neither acoustic nor elastic modeling methods can provide accurate simulations of waveforms. Besides, the point-wise comparison of wave-type signals is not good in general (Van Leeuwen and Mulder, 2010). As a result, phase- or traveltime-matching objective functions with extensions (in time or offset) are more practical. Crosscorrelation-type objective functions emphasize matching the phases of the predicted and the observed data, and thus, are insensitive to amplitude errors caused by the simplified physics we use to represent the medium (Choi and Alkhalifah, 2012). The non-zero lag crosscorrelation based FWI (Van Leeuwen and Mulder, 2008, 2010; Zhang et al., 2018a) have an additional degree
of freedom when comparing two data sets, which breaks the half-cycle difference criterion of the point-wise waveform comparison. However, the global crosscorrelation is often biased towards strong arrivals and is vulnerable to crosstalks between neighboring arrivals when using a large time lag. Isolating different arrivals and then comparing them separately can help us avoid these problems but it is often difficult and adds additional computational cost. A better solution is to use local crosscorrelation, which compares different arrivals locally using extensions instead of isolation (Díaz and Sava, 2015; Zhang and Alkhalifah, 2018).

We propose a local-similarity measurement to compare the observed and predicted data with variable time lags in elastic FWI. The variable time lags (an additional degree of freedom) are calculated from the local similarity (phase difference) of two traces with one of the two traces squeezed/stretched (Fomel and Jin, 2009). Instead of explicitly minimizing the time shift (Ma and Hale, 2013; Warner and Guasch, 2014; Zhu and Fomel, 2016; Bai et al., 2017), we maximize the local similarities of the two data. A weighted summation over time extensions avoids calculating/picking the matching filter (or the stationary curve) explicitly. The summation over all curves includes the primary stationary curve that plays a preferential role in other non-zero lag FWI implementations, but also multiple stationary curves when they exist (Landa et al., 2006). Besides, the local similarities with squeezing/stretching preserve the energy information of crosscorrelation, which can be an indicator of the minimum extension. A polynomial-type weighting function decided by the predefined extension is introduced to penalize localized crosscorrelations away from the stationary curve. It varies smoothly within the predefined extension and has fewer parameters to play around with than the widely used Gaussian or Linear ones.

This paper is divided into five sections. After the introduction, the theory of local-similarity based elastic FWI algorithm is described where we analyze the advantages of
using squeezing/stretching instead of shifting. We first use a simple perturbation model to analyze the sensitivity kernels obtained using a conventional $L_2$ norm and the proposed method. We then use a synthetic elastic model and the North Sea field data to verify the effectiveness of the proposed method. The sensitivity kernel analysis indicates that the proposed method can obtain a correct updating direction when the $L_2$ norm fails. Since this inversion corresponds to multiple parameters, the crosstalk between $v_p$ and $v_s$ always exists. The uncertainty analysis of the inverted parameters requires further investigations, which can be performed in future work. In the discussion, we compare the proposed approach with other algorithms such as the adaptive FWI.

**THEORY**

The objective of classical full waveform inversion is minimizing the $L_2$ norm of the difference between the observed and reproduced seismic data, which is written as

$$
\phi(m) = ||u(m, x_r, t) - d(m, x_r, t)||^2,
$$

(1)

where $\phi(m)$ measures the waveform differences; $u(m, x_r, t)$ and $d(m, x_r, t)$ are predicted and observed two-component data at $x_r$, respectively. $t$ indicates the recording time.

The first-order elastic wave equation is given by (Virieux, 1986; Vigh et al., 2014)

$$
\begin{pmatrix}
\rho I_3 & 0 \\
0 & C^{-1}
\end{pmatrix}
\frac{\partial \Psi(x, t)}{\partial t} - \begin{pmatrix}
0 & E^T \\
E & 0
\end{pmatrix} \Psi(x, t) - f(x_s, t) = 0,
$$

(2)

where $\Psi(x, t) = (v_1, v_2, v_3, \sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6)^T$ is a vector containing three particle velocities and six stresses, $I_3$ is a 3 by 3 identity matrix. $C$ is the stiffness matrix, $E$ denotes space differentiation, and $f(x_s, t)$ is the source located at $x_s$. 

7
Instead of minimizing the waveform difference, we propose to maximize the similarity between the observed and predicted seismic data using a local-crosscorrelation, which was used earlier in solving the time-lapse image registration problem (Fomel and Jin, 2009). The proposed objective function compares the predicted and observed data locally within the predefined stretching/squeezing range, which is given by

$$J(m) = -\frac{1}{2} \sum_s \sum_{x_r} \sum_t \sum_k W(k) C_w^2(m, s, x_r, t, k),$$  

where $C_w(m, s, x_r, t, k)$ denotes a localized crosscorrelation defined in equation 4. The two-components of the data are treated independently. The symbols $s$ and $x_r$ denote the source and receiver index numbers, respectively. $W$ is the weighting function. $k$ is the coefficient for data extension (stretching or squeezing).

The local-crosscorrelation admits time-dependent similarities instead of a single value often admitted by the global-crosscorrelation. The definition of local-crosscorrelation is given by

$$C_w(m, s, x_r, t, k) = \frac{\int_{t-rac{1}{2}w}^{t+rac{1}{2}w} u(m, s, x_r, t, t') d(m, s, x_r, t, k, t') dt'}{\sqrt{\int_{t-rac{1}{2}w}^{t+rac{1}{2}w} u^2 (m, s, x_r, t, t') dt'} \sqrt{\int_{t-rac{1}{2}w}^{t+rac{1}{2}w} d^2 (m, s, x_r, t, k, t') dt'}},$$  

where $u(m, s, x_r, t, t')$ and $d(m, s, x_r, t, k, t')$ are the modeled and extended observed data, respectively. $t'$ indicates the summation index within a time window, $w$; $k$ indicates the extension index of the observed data. In practice, the local-similarity is calculated by the solving a least-square problem with shaping regularizations (Fomel, 2007):

$$c_w = \frac{SU^T d}{\lambda^2 I + S (U^T U - \lambda^2 I)} \circ \frac{SD^T u}{\lambda^2 I + S (D^T D - \lambda^2 I)}.$$  

The local-similarity attributes are calculated by a product of two least-squares inverses using shaping regularization. The $S$ in equation 5 denotes the shaping operator, and $U$ and $D$ are diagonal matrices, which are composed of $u(x_r, t)$ and $d(x_r, t, kt)$, respectively.
I is the identity matrix and λ is the damping factor, and ⊙ denotes the component-wise product. The time window, \( w \), acts like the smoothing length in this case. We refer the reader to Fomel (2007) for more details.

We introduce a polynomial-type weighting function to replace the previously used linear or Gaussian functions, which satisfies the following boundary conditions: \( W|_{k_{\min}} = 0; W|_{k_{\max}} = 0; W|_1 = 1; W'|_{k_{\min}} = 0; W'|_{k_{\max}} = 0; W'|_1 = 0 \). Figure 1 shows one example of the weighting function. The Gaussian-type weighting functions requires another parameter, the standard deviation, to decide the shape of the penalty function. A narrow Gaussian function will remove most of the coherent arrivals and a wide Gaussian function cannot decay fast enough to reach zero at \( k_{\min} \) and \( k_{\max} \). The proposed weighting function does not need an additional parameter and it is guaranteed to be smooth within \([k_{\min}, k_{\max}]\). Also, the computation cost is bounded by the predefined maximum extension.

Conventional extensions usually compare two data sets by shifting one of them with respect to the other. In our proposed method, we use the stretching/squeezing approach to extend the observed data. The stretching/squeezing approach might be able to preserve the energy information compared to the shifting approach, which is more friendly to our proposed weighted-integration objective function. Smaller stretching/squeezing values tend to preserve more energy, which can be an indicator of the right matching direction from local-similarity panels. We use two 1D traces to demonstrate these assumptions numerically. Figure 2a shows the observed data with three arrivals, and the last arrival (E3) is supposed to match the predicted data in Figure 2b since they have the shortest distance. Two similarity panels calculated by shifting and stretching/squeezing are plotted in Figures 3a and 3b, respectively. Both panels indicate that the predicted data have two potential directions to match the observed data. However, there is only one correction direction which can be
found in the stretching/squeezing approach. In this local-similarity panel (Figure 3b), the energy (amplitudes) in two directions are different, the nearest matching (or the shortest distance) tends to have a higher amplitude, which is also the right direction to match the corresponding arrivals. The stationary curves (pink lines) picked from the panels are used in previous FWI algorithms (Ma and Hale, 2013; Zhu and Fomel, 2016; Bai et al., 2017). However, picking a stationary curve from the local-similarity panel will result in loss of information (i.e., amplitudes), which can be an indicator of correct extension direction. A weighted summation of the local similarities over extensions tends to find the available stationary curves without loss of information. The crosstalk between two arrivals can still exist in the proposed method as Figure 4 indicates. These two local similarity panels are local similarities of the observed data itself. The stretching/squeezing approach (Figure 4b) tends to have fewer local minima than the shifting approach (Figure 4a). Our proposed weighted integral approach can penalize the energy away from the expected stationary curve \((k = 1)\), and thus, can suppress local minima in the objective function. A picking approach can easily get trapped by one of the peaks away from the exact stationary curve in this case.

An elastic approximation to the Earth reduces the errors caused by the simplified physics often used in FWI, given by the acoustic approximation. However, estimating elastic parameters makes the inverse problem even more ill-posed. Regularization techniques can reduce the null space of the ill-posed inverse problem (Asnaashari et al., 2013; Wang et al., 2013; Duan and Sava, 2016; Zhang et al., 2018b). One efficient approach is using the smoothed total variation (TV) regularization (Alkhalifah et al., 2018), which is given by

\[
R_{TV}(m) = \int \sqrt{\epsilon^2 + \|
abla m\|^2} dx, \tag{6}
\]
where the gradient is given by \( g_m(R_{TV}) = \text{div} \left( \frac{\nabla m}{\sqrt{\epsilon^2 + ||\nabla m||^2}} \right) \), and \( \epsilon \) is a predefined parameter that makes equation 6 differentiable. A large \( \epsilon \) transforms the smoothed TV gradient to a scaled Laplacian (scaled by \( 1/\epsilon \)).

In this study, a TV regularization is applied in the field data test because of its effectiveness in suppressing noise. The exact TV regularization (\( \epsilon = 0 \) in equation 6) is implemented by an efficient primal-dual hybrid gradient algorithm (Zhu and Chan, 2008).

The gradient (for a source and receiver pair) of the misfit (objective) function in equation 3 can be calculated from the adjoint-state method (Plessix, 2006; Zhang et al., 2018a),

\[
g(m) = \left( \frac{\partial u(m,s,x_r,t)}{\partial m} \right)^T \int_k W(k) C_u(m,s,x_r,t,k) \sum_{i=1}^{6} \frac{d(m,s,x_r,t,k)}{\sqrt{\int_{t-\frac{1}{2}w}^{t+\frac{1}{2}w} u^2(m,s,x_r,t,t')dt'}} \sum_{j=1}^{6} \frac{u^2(m,s,x_r,t,t')dt'}{\sqrt{\int_{t-\frac{1}{2}w}^{t+\frac{1}{2}w} d^2(m,s,x_r,t,k,t')dt'}} \right) dk. \tag{7}
\]

The local crosscorrelation terms are calculated using equation 5, while the remaining denominator terms are approximated using a full-time integral, which is a reasonable approximation and it also helps stabilize the division.

For the parameterization of \( C_{ij} \), the Fréchet derivative, \( \frac{\partial u(C_{ij},x,t)}{\partial C_{ij}} \), is given by Vigh et al. (2014):

\[
\frac{\partial u(C_{ij},s,x_r,t)}{\partial C_{ij}} = \frac{\partial C}{\partial C_{ij}} C^{-1} \left( \frac{\partial \sigma}{\partial t} - f \right)_{i=1,...,6;j=i,...,6} \quad \text{and} \quad \left( \frac{\partial C}{\partial C_{ij}} \right)_{pq} = \begin{cases} 
1, & p = i, q = j \\
1, & p = j, q = i \\
0, & \text{otherwise}
\end{cases}
\tag{8}
\]

where \( \sigma \) denotes the stress component of the forward-propagated wavefield. \( \frac{\partial C}{\partial C_{ij}} \) is a six-by-six matrix with elements defined in equation 8. Here we use the parameterization of \( C_{ij} \),
but the gradients for other parameters such as \( v_p \) and \( v_s \) can be derived by using the chain rule.

The model is updated iteratively using the lBFGS method (Liu and Nocedal, 1989), which is written as

\[
m = m_0 - \alpha H^{-1}g,
\]

where \( \alpha \) is the step length calculated by satisfying the Wolfe condition (Wolfe, 1969), and \( H \) is the approximated Hessian matrix.

**NUMERICAL EXAMPLES**

**Sensitivity kernel analysis**

We first analyze the sensitivity kernels obtained using the objective functions of the conventional \( L_2 \) norm and the proposed objective function for one source-and-receiver pair.

The actual model has a homogeneous background and two elliptical anomalies in \( v_p \) and \( v_s \), respectively, as shown in Figure 5. We use constant background velocities as initial models (Figures 5c and 5d). The solid star and triangle indicate the locations of source and receiver, respectively. The P- and S-wave velocity anomalies, which indicate the required updating directions, are plotted in Figures 6a and 6b, respectively. The source has a peak frequency of 8 Hz, in which case the conventional \( L_2 \) norm FWI cannot provide us with the correct updating directions as shown in Figures 6c and 6d. The first Fresnel zones indicate opposite updating directions compared to the desired updates. However, the sensitivity kernels of the proposed method as shown in Figures 6e and 6f have the right updating directions. A larger stretching/squeezing range allows a larger search area when comparing the observed and predicted data. Zhang et al. (2018a) suggest using the auto-correlation
of observed data to decide a rough extension range, which also could be useful here. In this case, we use a range of 0.8-1.2 to achieve a proper comparison and we obtained similar updates when using larger extensions. Although the velocity anomalies in \(v_p\) and \(v_s\) are in different horizontal locations, the calculated updates are smeared along the wave-path as expected.

**The modified Elastic Marmousi Model**

We use a modified Marmousi model to verify the effectiveness of the proposed objective function. Two low-velocity zones (indicated by the dashed circles in Figure 7a) in the P-wave velocity are intended to test whether there are leakages between the inverted \(v_p\) and \(v_s\). The actual S-wave velocity (Figure 7b) is generated by setting \(v_s = \frac{v_p}{\sqrt{3}} + 0.1(v_p - 2.4)\). Initial models in Figure 8 are 1D linear gradient models, which are far from the exact ones \((v_s = \frac{v_p}{\sqrt{3}})\). The source wavelet is a Ricker wavelet (peak frequency, \(f_p = 5\) Hz) without frequencies below 3 Hz, in which case a conventional zero-lag crosscorrelation based elastic FWI is supposed to converge to a local minimum. Only one frequency band, 3-10 Hz, is used for inversion. The stretching/squeezing range is fixed, which is 0.8-1.2. There are 220 explosive sources and 330 receivers placed at constant increments on the surface of the model. The recorded data are two-component particle velocities. A fourth-order in space and second-order in time staggered finite-difference scheme with optimal finite-difference coefficients is implemented to solve the elastic wave equation (Virieux, 1986; Liu, 2014). The spatial and time sampling are 10 m and 1 ms, respectively. A boundary-storing approach is used in wavefield reconstruction (Raknes and Weibull, 2016). We update \(v_p\) and \(v_s\) simultaneously using the LBFGS algorithm. 20 gradients are used to approximate the Hessian matrix. We conduct a zero-lag global-correlation based FWI and the proposed FWI
on the same data set for comparison. Figure 9 shows the estimated velocities using a zero-lag global correlation. They’re not updated sufficiently because of the convergence to a local minimum. Figure 10 shows the estimated velocities using the proposed method. Both $v_p$ and $v_s$ got improved compared to those obtained from the global correlation objective. We plot the vertical profiles in Figure 11 for a better comparison. The estimated velocities using the proposed method have better match with the actual ones than those estimated using the global correlation approach. As a quality control measure, we plot the local-similarity panels for the initial (Figure 12a), inverted (Figure 12b) and actual (Figure 12c) models. Positive values indicate that the two data sets are close to each other while the negative values indicate that the two data sets are more than half-cycle away. Local-similarities between the predicted data from the inverted model and the observed data are concentrated near $k = 1$, which means that the predicted is close to the observed data. From Figure 12, we can understand the optimizing mechanism of the proposed method: it corrects the moveouts of the arrivals in data domain, which is similar to the process of moveout velocity analysis in the image domain.

At last, we use the estimated velocities from the proposed method as initials for a conventional $L_2$ norm based FWI and obtain the results in Figure 13. The final result provides models that are close to the actual ones and two low-velocity zones have been recovered. The proposed method can produce good initial models for conventional $L_2$ norm based inversion as Figure 14 indicates. The data matching history shown in Figure 15 indicates that the predicted data from the inverted model is also close to the observed data.
The North Sea OBC data

We apply the proposed inversion strategy to a 2D line from the Volve data set. The seal layer and the reservoir, located between 2.75-3.12 km depth, are the main imaging targets. We use the raw data set with limited preprocessing including polarity correction, instrumental deconvolution and data quality control. For the inversion, we use 240 shots and 240 two-component receivers distributed evenly with a spacing interval of 50 m and 25 m, respectively. The length of the ocean-bottom cable (OBC) is 6 km and the sources are evenly distributed along a 12 km line just below the sea surface. A modified free-surface boundary condition, which can suppress strong surface waves, is used in the simulation (He et al., 2016). We convolve the observed data with a half-order differentiation function of the known wavelet, and thus, we avoid source estimation, while correcting the phase and amplitude discrepancy between the 3D Earth and the 2D simulation (Pica et al., 1990; Yoon et al., 2012). The initial model (Figure 16) is a 1D smoothed version of the model provided by the data owners. Only one frequency band (2-12 Hz) is used for the inversion. The relatively high frequencies help suppress the Scholte waves present in the shallow-water area. We first apply our proposed inversion algorithm with an extension of \( k = 0.9 - 1.1 \) to avoid possible cycle-skips. The inverted \( v_p \) and \( v_s \) are shown in Figures 17a and 17b, respectively. The updates are dominated in the shallow area, which indicates that the contribution is mainly from the first arrivals. There are also minor updates from the reflections, which can be found through the data comparison. Then we refine the model obtained in the previous stage using a zero extension (\( k = 1.0 \)), in which case the proposed method is equivalent to the global-crosscorrelation approach (Choi and Alkhalifah, 2012). Figure 18 shows the final inverted velocities. The high-velocity seal layer and a low-velocity layer below it can be imaged. For comparison, we also apply an \( L_2 \) norm based inversion and obtained minor updates in the
shallow area as shown in Figure 19. There are no sufficient updates in \( v_p \) and \( v_s \) compared to the proposed approach. Also, considering the limited offsets available, the inverted \( v_s \) has larger uncertainties than \( v_p \). For a fair comparison, the inversion scenarios share the same parameter setting with regard to gradient post-processing (smoothing and model regularization), optimization strategy and stopping criteria (the Wolfe condition) (Wolfe, 1969). We also compare the inverted P-wave velocities with the one from the check shot nearby in Figure 20. The large extension range \((k = 0.9 - 1.1)\) provides relatively smooth updates compared to those with non-extension \((k = 1.0)\). The extension helps correct the kinematic errors in the initial model. Since anisotropy is present in this data (Guitton and Alkhalifah, 2017), the inverted velocity here is closer to the horizontal velocity in the shallow and the NMO velocity in the deeper part. On the other hand, the well velocity represents the vertical velocity. This explains the discrepancy in the matching, where the difference can be attributed to \( \delta \). This also explains the slight depth mismatch as the inverted velocity will place features slightly deeper. The interleaved predicted and observed vertical- and horizontal-component data plot (Figure 21) of the initial model indicates that the initial model can provide reasonably good prediction in the near offsets. The inversion with the extension helps us match the data at the far offsets and recovers some dominant reflections as shown in Figure 22. The interleaved data comparisons indicate that the reduced extension helps us obtain simulated data that match the observed data even better (Figures 22 and 23). As a reference, the \( L_2 \) based inversion failed to provide sufficient updates (satisfying the Wolfe condition) as shown in Figure 24. We also plot the data-matching history for the different inversion scenarios as shown in Figure 25. It shows that the proposed inversion reduces the data misfit by about 45% while the conventional \( L_2 \) norm achieves about 10% reduction of data misfit.
DISCUSSION

The proposed method can be regarded as a data-domain analysis approach. It tries to focus the local-similarities along the non-extension range as shown in Figure 12. The kinematic information, which is captured by the extension, is mainly utilized when calculating the updates. The squeezing/stretching extends the search dimension when comparing two data sets and thus allows us to start from kinematically poor initial velocity models. The sensitivity kernel analysis (Figure 6) demonstrates that the proposed method can find the correct updating direction when the conventional $L_2$ norm fails. As opposed to dynamic wrapping (Ma and Hale, 2013), we did not calculate explicit traveltime differences. The penalty function used in the objective function tries to minimize the time difference implicitly. Unlike methods that rely on utilizing a matching filter (Warner and Guasch, 2014; Zhu and Fomel, 2016), which requires an initial deconvolution, we use a weighted summation of filters to obtain a stationary filter (Landa et al., 2006). These filters are represented by the local-similarities of predicted and observed data, which emphasize the matching of the phases instead of the amplitudes. In practice, artifacts can bias the calculation of one stationary filter and the weighted summation can increase the robustness of the algorithm.

A potential limitation of the proposed method, as many other measurements do, is that it requires diving waves to provide sufficient low-wavenumber components for kinematically poor initial velocity models. A remedy to this is to calculate updates along reflection wave-paths explicitly. Our proposed objective function can be used as an objective to emphasize matching of phase information somewhat similar to what traveltime based reflection FWI does (Wang et al., 2017; Wang and Cheng, 2017).

Multiparameter estimation increases the ill-posedness of the inverse problem. Regular-
ization, such as total-variation, can effectively reduce the number of unknowns and thus reduces the nonuniqueness. Besides, total-variation can help in extending the range of model wavenumbers in updates (Alkhalifah et al., 2018). As an alternative, utilizing different geophysical data can also help to resolve multiparameters better. The interpretation of seismic data on their own will provide incomplete information due to the non-uniqueness and the limited spatial resolution. However, additional measurements that may illuminate the reservoir with varying coverage and resolution can provide considerable value (Hu et al., 2009). The way to suppress the crosstalks between multiple parameters using a hierarchical inversion strategy and using facies constraints have been discussed by Oh and Alkhalifah (2016) and Zhang et al. (2018b), respectively, which is beyond the scope of this paper.

Our proposed method has a similar computational cost to conventional FWI if we use a relatively large smoothing window and fewer iterations when solving equation 5. With an appropriate choice of a smoothing window and one iteration, equation 5, proposed by Fomel (2007), is equivalent to the fast local-crosscorrelation algorithm proposed by Hale (2006), which is computationally efficient. It can be easily applied to 3D acquisition since our comparisons are done trace-wise and do not depend on the acquisition geometry. The proposed method can be easily incorporated into the conventional $L_2$ norm based FWI, by a simple modification to the objective function and the corresponding adjoint source. The optimization problem can be solved by current gradient-based optimization algorithms such as the lBFGS method.

**CONCLUSIONS**

The widely used wiggle-to-wiggle subtraction fails to provide the correct direction for model updating when the signals differ in phase by more than a half-cycle. It also fails to do so
when the signals have significant differences in seismic signature, which can arise due to attenuation. The proposed similarity-based full waveform inversion algorithm intends to maximize the local similarity of observed and predicted data. With an extension, the stretching/squeezing, to the data set, we can compare two arrivals that are otherwise cycle-skipped. The stretching/squeezing information is utilized to boost up the low-wavenumber components of the model. Instead of picking a stationary curve from the local-similarity panel, we use a weighted integral to find the possible stationary curves. It avoids unstable and tedious picking and allows us to find multiple stationary curves when they exist. The local similarity can include crosstalk between neighboring arrivals in both the stretching/squeezing or shifting approaches. However, the stretching/squeezing can preserve more energy information compared to the shifting approach. With the weighting function applied to the similarity panels, crosstalks away from the stationary curve can be suppressed. Both synthetic and field examples verify the effectiveness of the proposed objective function. The field example also shows that the intended objective function is relatively immune to noise recorded in the shallow-water area. However, it does not aim to solve the crosstalks prominent in multiple parameter inversion. Compared to the global-correlation measurement, the proposed method can have a similar computational cost to calculate local similarities.

ACKNOWLEDGMENTS

We thank Yike Liu and Zedong Wu for their helpful discussions. We want to thank Yang Liu, Wiktor Weibull and two anonymous reviewers, for the effort put into the review of this paper. We thank KAUST for its support and specifically the seismic wave analysis group members for their valuable insights. We thank Equinor and the former Volve license partners ExxonMobil E & P Norway AS and Bayergas Norge AS, for the release of the
Volve data. The views expressed in this paper are the views of the authors and do not necessarily reflect the views of Equinor and the former Volve field license partners. For computer time, this research used the resources of the Supercomputing Laboratory at King Abdullah University of Science & Technology (KAUST) in Thuwal, Saudi Arabia.
REFERENCES


Routh, P., J. Krebs, S. Lazaratos, A. Baumstein, I. Chikichev, S. Lee, N. Downey, D.


Yang, Y., B. Engquist, J. Sun, and B. F. Hamfeldt, 2018, Application of optimal transport


———, 2019a, Regularized elastic full waveform inversion using deep learning: Geophysics, 84, 1–47.


LIST OF FIGURES

1 The proposed weighting function. It is determined by the predefined maximum $k$ and reduces to zero at the extension boundaries.

2 One simple trace for illustration. a) Observed data and b) predicted data. The arrival in the predicted data (e3) is supposed to match the arrival E3 in the observed data.

3 Local similarities of the predicted and observed trace. a) The shifting approach and b) the stretching/squeezing approach. Pink curves are picked from the similarity panels (Fomel and Jin, 2009). Local-crosscorrelation tries to match the observed arrivals, E1 and E3, to the predicted arrival, e3, with different directions of shifting or stretching/squeezing. There is a significant amplitudes difference in the stretching/squeezing approach. Fewer stretching/squeezing tends to preserve larger amplitudes (energy).

4 Local similarities of the observed trace. a) The shifting approach and b) the stretching/squeezing approach. In an ideal case, the energy should focus on the central line. However, due to the crosstalks between neighboring arrivals, there is some energy away from the central line. This artificial energy (indicated by the white arrows) is harmful in extracting a stationary filter from such panels. The proposed weighted summation approach will penalize these crosstalks in the inversion.

5 Velocity models. a) Actual $v_p$, b) actual $v_s$, c) initial $v_p$ and d) initial $v_s$. The solid star and triangle indicate the source and receiver, respectively.

6 Sensitivity kernels. Actual $v_p$ and $v_s$ anomalies in a) and b), respectively. Sensitive kernels for $v_p$ and $v_s$ using the conventional $L_2$ measurement are shown in c) and d), respectively. Sensitive kernels for $v_p$ and $v_s$ using the proposed measurement are shown in
e) and f), respectively. The proposed method can find the correct updating direction while the conventional $L_2$ norm fails to find the correct updating direction.

7 The true model. a) $v_p$, b) $v_s$. There are two low-velocity zones marked by dashed circles in $v_p$; $v_s = v_p / \sqrt{3} + 0.1(v_p - 2.4)$.

8 The initial model. a) $v_p$, b) $v_s$. They’re linear gradient models and $v_s = v_p / \sqrt{3}$.

9 The estimated models using a conventional global crosscorrelation. a) $v_p$, b) $v_s$. We used limited iterations as the model converged to a local minimum.

10 The estimated model using the proposed approach. a) $v_p$, b) $v_s$. They’re used as part of the initial model for a conventional $L_2$ norm based elastic FWI.

11 Vertical profiles located in the middle of the models. a) $v_p$, b) $v_s$. The proposed estimates are close to the actual ones.

12 A local-similarity comparison (vertical component) for one example shot. a) Initial and observed data, b) inverted and observed data and c) auto-correlation of observed data. The inverted similarities focus on $k = 1$ as shown in b) and they are close to the reference in c).

13 Conventional $L_2$ norm based FWI using the velocities from the proposed method. a) $v_p$, b) $v_s$. Two low-velocity zones have been recovered in the estimated $v_p$ as the pink arrows indicate. Only one frequency band (3-10 Hz) is used in this inversion.

14 Vertical profiles located in the middle of the models. a) $v_p$, b) $v_s$. The initial model for the $L_2$ norm based FWI comes from the proposed method.

15 Normalized data misfit versus iterations for the conventional $L_2$ norm inversion. We use the model in Figure 10 as the initial.
The initial models. a) $v_p$ and b) $v_s$. They’re 1D models.

The inverted models using $k = 0.9 - 1.1$. a) $v_p$ and b) $v_s$. They are used as input models for the following non-extension inversion.

The inverted models using $k = 1.0$. a) $v_p$ and b) $v_s$. The high-velocity seal layers at the depth of 2.8 km are recoverable but the uncertainties in the inverted $v_s$ might be large due to the limited offset/depth ratio.

The inverted models from using the conventional $L_2$ norm as a reference. a) $v_p$ and b) $v_s$. The inversion was terminated when the updates cannot satisfy the Wolfe condition.

P-wave velocity profiles at $X = 6.0$ km. The check shot is close to the inversion line. The depth mismatch could be caused by the dipping layers or more likely the ignored anisotropy in this area.

A shot gather displaying interleaved predicted and observed data using the initial $v_p$ and $v_s$. a) The vertical component and b) the horizontal component.

A shot gather displaying interleaved predicted and observed data using the estimated $v_p$ and $v_s$ from $k = 0.9 - 1.1$. a) The vertical component and b) the horizontal component.

A shot gather displaying interleaved predicted and observed data using the estimated $v_p$ and $v_s$ from $k = 1.0$. a) The vertical component and b) the horizontal component.

A shot gather displaying interleaved predicted and observed data using conventional $L_2$ norm based inversion. a) The vertical component and b) the horizontal component.

Data matching history. There are two sequential inversions: 1) $k = 0.9 - 1.1$ and 2) $k = 1.0$. The conventional $L_2$ norm based inversion acts a reference. All these inversions
are terminated via the Wolfe condition.
Figure 1: The proposed weighting function. It is determined by the predefined maximum $k$ and reduces to zero at the extension boundaries.
Figure 2: One simple trace for illustration. a) Observed data and b) predicted data. The arrival in the predicted data (e3) is supposed to match the arrival E3 in the observed data.
Figure 3: Local similarities of the predicted and observed trace. a) The shifting approach and b) the stretching/squeezing approach. Pink curves are picked from the similarity panels (Fomel and Jin, 2009). Local-crosscorrelation tries to match the observed arrivals, E1 and E3, to the predicted arrival, e3, with different directions of shifting or stretching/squeezing. There is a significant amplitudes difference in the stretching/squeezing approach. Fewer stretching/squeezing tends to preserve larger amplitudes (energy).
Figure 4: Local similarities of the observed trace. a) The shifting approach and b) the stretching/squeezing approach. In an ideal case, the energy should focus on the central line. However, due to the crosstalks between neighboring arrivals, there is some energy away from the central line. This artificial energy (indicated by the white arrows) is harmful in extracting a stationary filter from such panels. The proposed weighted summation approach will penalize these crosstalks in the inversion.
Figure 5: Velocity models. a) Actual $v_p$, b) actual $v_s$, c) initial $v_p$ and d) initial $v_s$. The solid star and triangle indicate the source and receiver, respectively.
Figure 6: Sensitivity kernels. Actual $v_p$ and $v_s$ anomalies in a) and b), respectively. Sensitive kernels for $v_p$ and $v_s$ using the conventional $L_2$ measurement are shown in c) and d), respectively. Sensitive kernels for $v_p$ and $v_s$ using the proposed measurement are shown in e) and f), respectively. The proposed method can find the correct updating direction while the conventional $L_2$ norm fails to find the correct updating direction.
Figure 7: The true model. a) $v_p$, b) $v_s$. There are two low-velocity zones marked by dashed circles in $v_p$; $v_s = v_p/\sqrt{3} + 0.1(v_p - 2.4)$. 
Figure 8: The initial model. a) $v_p$, b) $v_s$. They’re linear gradient models and $v_s = v_p / \sqrt{3}$. 
Figure 9: The estimated models using a conventional global crosscorrelation. a) $v_p$, b) $v_s$.

We used limited iterations as the model converged to a local minimum.
Figure 10: The estimated model using the proposed approach. a) $v_p$, b) $v_s$. They’re used as part of the initial model for a conventional $L_2$ norm based elastic FWI.
Figure 11: Vertical profiles located in the middle of the models. a) $v_p$, b) $v_s$. The proposed estimates are close to the actual ones.
Figure 12: A local-similarity comparison (vertical component) for one example shot.  a) Initial and observed data, b) inverted and observed data and c) auto-correlation of observed data. The inverted similarities focus on $k = 1$ as shown in b) and they are close to the reference in c).
Figure 13: Conventional $L_2$ norm based FWI using the velocities from the proposed method.

a) $v_p$, b) $v_s$. Two low-velocity zones have been recovered in the estimated $v_p$ as the pink arrows indicate. Only one frequency band (3-10 Hz) is used in this inversion.
Figure 14: Vertical profiles located in the middle of the models. a) $v_p$, b) $v_s$. The initial model for the $L_2$ norm based FWI comes from the proposed method.
Figure 15: Normalized data misfit versus iterations for the conventional $L_2$ norm inversion.

We use the model in Figure 10 as the initial.
Figure 16: The initial models. a) $v_p$ and b) $v_s$. They’re 1D models.
Figure 17: The inverted models using $k = 0.9 - 1.1$. a) $v_p$ and b) $v_s$. They are used as input models for the following non-extension inversion.
Figure 18: The inverted models using $k = 1.0$. a) $v_p$ and b) $v_s$. The high-velocity seal layers at the depth of 2.8 km are recoverable but the uncertainties in the inverted $v_s$ might be large due to the limited offset/depth ratio.
Figure 19: The inverted models from using the conventional $L_2$ norm as a reference. a) $v_p$ and b) $v_s$. The inversion was terminated when the updates cannot satisfy the Wolfe condition.
Figure 20: P-wave velocity profiles at \( X = 6.0 \, km \). The check shot is close to the inversion line. The depth mismatch could be caused by dipping layers or more likely the ignored anisotropy in this area.
Figure 21: A shot gather displaying interleaved predicted and observed data using the initial $v_p$ and $v_s$. a) The vertical component and b) the horizontal component.
Figure 22: A shot gather displaying interleaved predicted and observed data using the estimated $v_p$ and $v_s$ from $k = 0.9 - 1.1$. a) The vertical component and b) the horizontal component.
Figure 23: A shot gather displaying interleaved predicted and observed data using the estimated $v_p$ and $v_s$ from $k = 1.0$. a) The vertical component and b) the horizontal component.
Figure 24: A shot gather displaying interleaved predicted and observed data using conventional $L_2$ norm based inversion. a) The vertical component and b) the horizontal component.
Figure 25: Data matching history. There are two sequential inversions: 1) $k = 0.9 - 1.1$ and 2) $k = 1.0$. The conventional $L_2$ norm based inversion acts a reference. All these inversions are terminated via the Wolfe condition.