The objective functions and adjoint sources behavior for elastic full waveform inversion

Tianze Zhang, Kristopher Innanen

ABSTRACT

Elastic full-waveform inversion is an ill-posed data-fitting procedure that is sensitive to noise, inaccuracies of the starting model, the definition of multi-parameter classes, and inaccurate modeling of wave fields amplitudes. The objective function measures the difference between the synthetic data and observed data, which plays an essential role in the convergence property of the elastic wave equation. In this study, we investigate the behavior of the ℓ_2, ℓ_1 norm and the correlation-based objective function in FWI, and show the sensitivity of these objective functions with respect to VP, VS, and density. The objective functions show that all three objective functions show strong non-linearity with the VS parameter, which means the inversion is relatively harder to invert compared with the other two parameters since it contains a lot of local minimums. We also developed an objective function based on the multi-scale Z transform. The multi-scale Z transform objective function could release the non-linearity of the objective function since the low frequency information of the objective function is first extracted. Thus, the adjoint source contains low frequency information, which helps to build the larger-scale information of the velocity model.

INTRODUCTION

FWI potentially provides high-resolution models of the subsurface, but it suffers from two main difficulties. However, the ill-posedness and the non-linearity of the inverse problem is one of the most important issues faced with full waveform inversion. The ill-posedness of FWI mainly arises from the lack of low frequencies in the source bandwidth and the incomplete illumination of the subsurface provided by conventional seismic surveys (Brossier et al., 2010). Thus, FWI is highly nonlinear, and the results strongly depend on the accuracy of the starting model in the framework of local optimization and on the presence of noise. Several hierarchical multi-scale strategies that proceed from low frequencies to higher frequencies have been proposed to mitigate the non-linearity of the inverse problem. Bunks et al. (1995) apply the multi-grid method to a subsampled, low frequency version of the Marmousi data set. Sirgue and Pratt (2004) have defined a strategy for selecting temporal frequencies for efficient prestack imaging. Brossier et al. (2009) suggests that simultaneous inversion of multiple frequencies is critical when considering complex wave phenomena. Wu et al. (2014) recognized that the envelope fluctuation and decay of seismic records carries ultra low-frequency (i.e., the frequency below the lowest frequency in the source spectrum) signals that can be used to estimate the long-wavelength velocity structure. We then developed envelope inversion for the recovery of low wavenumber components of media (smooth background), so that the initial model dependence of waveform inversion can be reduced.

The study of the objective functions for FWI could help us to understand the non-linearity of FWI. The least-squares objective function remains the most commonly used criterion in FWI, although it theoretically suffers from poor robustness in the presence of large isolated and non-Gaussian errors Brossier et al. (2009). The ℓ_1 norm is not based on the Gaussian statistics in data space, and it was first introduced in the time domain with Tarantola (2005). Djikpéssé and Tarantola (1999) used the ℓ_1 norm to successfully invert the field data in the Gulf of Mexico. Pyun et al. (2009) use the ℓ_1 norm on the real and imaginary part of the frequency domain components and achieve promising results. Huber norm (Huber, 1973) or the hybrid ℓ_2 and ℓ_1 norm combines the advantages of the ℓ_1 and ℓ_2 norm to release the non-linearity of the objective function. The small residuals would be updated according to the ℓ_2 norm, and the large residuals would be updated with the ℓ_1 norm. Therefore, the hybrid objective function has the ability to tolerate the outliers noise in data.

The phase information contains the kinematics property of wave propagation. Thus it has great potential to be used in FWI. The cycle skipping problem could be explained from the phase aspects of the wave fields. The phase of the Fourier-transformed wave field, or what is referred to as the principal value of phase (PVP), has a finite but periodic range [-π, π] thus phase of the modeled data could converge to the wrong phase corresponding to a different cycle in the procedure of waveform inversion, which makes the inversion converge to a local minimum. Thus, the utilization of the phase in the objective function also raises the interest of the researchers. Liu et al. (2017), introduced the full waveform inversion with normalized zero-lag cross-correlation objective function to relax on the amplitude constraints and emphasizes the phase information when measuring the closeness between the simulated and observed data. This FWI method becomes insensitive to differences in amplitude. They test the effectiveness and robustness of FWI with the normalized zero-lag cross-correlation function against the noise and unpredictable amplitude of the data that cannot be modeled by the wave fields extrapolation operator. The direct utilization of using the phase in the objective function suffers from the phase wrapping problem, which limits its applications. Some approaches have been proposed to overcome this problem. Choi and Alkhalifah (2015) introduced the instantaneous travel time in the inversion, which is obtained from the phase derivative with respect to the angular frequency. Shah et al. (2012) proposed the phase-unwrapped full waveform inversion (FWI) methodology for applying the technique to seismic data directly from a poor or simple starting model in an automated, robust manner.

In this report, we first take a brief review of the objective functions and their adjoint sources. We review the ℓ_2, ℓ_1 and correlation-based objective functions. We discuss how these objective functions react to the variation of VP, VS, and density in velocity parameterization. The contour plot of the objective functions demonstrates the non-linearity of these objective functions and shows the local minimums of the objective functions. We also show that the variation of VS parameter would cause severe non-linearity of the objective function, which could make VS hard to invert and converge. We also design a multi-scale Z transform objective functions, which measures the different damping properties in the shotrecords with a scale factor z, which makes the inversion to invert different waveform components in the shot records, for instance, from surface waves to reflection waves. The contour plot of the multi-scale Z transform objective functions is much smoother than the other three objective functions and has few local minimums. We also do the frequency analysis of the adjoint sources for the multi-scale Z transform objective function. It shows that its adjoint sources have abundant low frequency information, which could help the inversion to build large-scale information of the velocity model and reduce the non-linearity of the inversion.

Least-squares norm

The Least-squares norm is often referred as the ℓ_2 norm. The ℓ_2 norm objective function is one of the most commonly used objective functions in FWI and is expressed as:

$$E(m)_{\ell_2} = \frac{1}{2}||d_{cal}(r,t,m) - d_{obs}(r,t)||^2 \tag{1}$$

The $d_{cal}(r,t,m)$ is the synthetic data which is recorded on receiver r, at time t and is parameterized with parameter m. $d_{obs}(r,t)$ is the observed data recorded on receiver r, at time t. $E(m)$ is the objective function that parameterized with m. Differentiating $E(m)$ with respect to the model parameters gives the adjoint source $\mathcal{X}_{\ell_2} = d_{cal}(r,t,m) - d_{obs}(r,t)$. We could use this adjoint source to calculate the back-propagation and generating the back-propagation wave fields. Gradients for the parameters could be gained with:

$$\frac{\partial E(m)}{\partial m} = \sum_r \int_t \frac{\partial d_{obs}(r,t,m)}{\partial m} \mathcal{X}_{\ell_2} \, dt \tag{2}$$

, where the $\frac{\partial d_{obs}(r,t,m)}{\partial m}$ is the Fréchet derivative of synthetic data with respect to the parameter m In 2, the $d_{cal}(r,t,m)$ is the synthetic data and the $d_{obs}(r,t)$ is the observed data. m is the model parameter.

Least-absolute norm

The least absolute norm is also referred as the ℓ_1 norm objective function, introduced with (Tarantola, 2005; Crase et al., 1990) which can be expressed as:

$$E(m)_{\ell_1} = ||d_{cal}(r,t,m) - d_{obs}(r,t)|| \tag{3}$$

Equation 3 shows that the ℓ_1 norm is the absolute error between the synthetic data and the observed data. The ℓ_1 norm is not analytically differentiable. However the adjoint source of the ℓ_1 norm is:

$$\mathcal{X}_{\ell_1} = \frac{d_{cal}(r,t,m) - d_{obs}(r,t)}{||d_{cal}(r,t,m) - d_{obs}(r,t)||} \tag{4}$$

Equation 4 can also be expressed with the use of sign function Tarantola, 2005. According to Brossier et al. (2010) in ℓ_1 norm adjoint source, the data residuals are normalized according to their amplitudes, which gives insight into why this is expected to be less sensitive to large residuals.

Correlation-based objective function

The global correlation function can be expressed as:

$$E(m)_{GC} = -\sum\nolimits_r \frac{\int_t \left[d_{cal}(r,t,m) d_{obs}(r,t)\right] dt}{\sqrt{\int_t \left[d_{cal}(r,t,m)\right]^2 dt} \sqrt{\int_t \left[d_{obs}(r,t)\right]^2 dt}} \tag{5}$$

From equation 5, we can see that the global correlation is defined by the zero-lag cross-correlation of two normalized signals Choi and Alkhalifah (2012). One of the motivations for introducing the correlation-based objective function is to match more phase information into the inversion. Since solely matching the amplitude could have the problem of trapping into the cycle skipping problem. Equation 6 shows the adjoint source of the global correlation objective function.

$$\mathcal{X}_{GC} = \frac{\int_t [d_{cal}(r,t,m)][d_{obs}(r,t)]\, dt d_{cal}(r,t,m)}{\sqrt{\int_t [d_{cal}(r,t,m)]^2\, dt}^3 \sqrt{\int_t [d_{obs}(r,t)]^2\, dt}} - \frac{d_{obs}(r,t)}{\sqrt{\int_t [d_{cal}(r,t,m)]^2\, dt}\sqrt{\int_t [d_{obs}(r,t)]^2\, dt}} \tag{6}$$

Another correlation based objective functions is the zero mean global correlation misfit function, which is expressed as:

$$E(m) = -\sum\nolimits_r \frac{\int_t \left[d_{cal}(r,t,m) - \overline{d}_{cal}(r,t,m)\right]\left[d_{obs}(r,t) - \overline{d}_{obs}(r,t)\right] dt}{\sqrt{\int_t \left[d_{cal}(r,t,m) - \overline{d}_{cal}(r,t,m)\right]^2 dt}\sqrt{\int_t \left[d_{obs}(r,t) - \overline{d}_{obs}(r,t)\right]^2 dt}} \tag{7}$$

where $\overline{d}_{cal}(r,t,m)$ are the mean value of for the synthetic data trace and $\overline{d}_{obs}(r,t)$ is the mean of the observed data trace. The adjoint source for this objective function is:

$$\mathcal{X}_{zeroGC} = \frac{1}{\sqrt{\int_t \left[d_{cal}(r,t,m) - \overline{d}_{cal}(r,t,m)\right]^2 dt}} (\widetilde{d_{cal}} \widetilde{d_{cal} d_{obs}} - \widetilde{d_{obs}}) \tag{8}$$

$\widetilde{d_{cal} d_{obs}}$, $\widetilde{d_{cal}}$, $\widetilde{d_{obs}}$ has the expression of:

$$\widetilde{d_{cal} d_{obs}} = \frac{\int_t \left[d_{cal}(r,t,m) - \overline{d}_{cal}(r,t,m)\right]\left[d_{obs}(r,t) - \overline{d}_{obs}(r,t)\right] dt}{\sqrt{\int_t \left[d_{cal}(r,t,m) - \overline{d}_{cal}(r,t,m)\right]^2 dt}\sqrt{\int_t \left[d_{obs}(r,t) - \overline{d}_{obs}(r,t)\right]^2 dt}} \frac{d_{cal}(r,t,m) - \overline{d}_{cal}(r,t,m)}{\sqrt{\int_t \left[d_{cal}(r,t,m) - \overline{d}_{cal}(r,t,m)\right]^2 dt}} \tag{9}$$

$$\widetilde{d_{cal}} = \frac{[d_{cal}(r,t,m) - \overline{d}_{cal}(r,t,m)]}{\sqrt{\int_t \left[d_{cal}(r,t,m) - \overline{d}_{cal}(r,t,m)\right]^2 dt}} \tag{10}$$

$$\widetilde{d_{obs}} = \frac{[d_{obs}(r,t,m) - \overline{d}_{obs}(r,t,m)]}{\sqrt{\int_t \left[d_{obs}(r,t,m) - \overline{d}_{obs}(r,t,m)\right]^2 dt}} \tag{11}$$

The zero mean global correlation objective function performs well when there is random noise in the data Dong et al. (2019).

Equations 2, 4, 6, 7 show that the gradients of the objective functions have similar forms but different source terms for the back-propagated adjoint wave fields. This implies that the same FWI algorithm can be used to compute the gradients of the different misfit functions with the same computational cost, provided the source term of the adjoint back-propagated wave fields and the misfit function can be computed for each functional.

Multi-scale Z transform objective function

Here we introduce the The multi-scale Z transform objective function which is:

$$E(m)_{MZ} = \frac{1}{2}||\underline{F_z} d_{cal}(r,t,m) - \underline{F_z} d_{obs}(r,t)||^2 \tag{12}$$

,where the matrix $\underline{F_z}$ is:

$$\underline{F_z} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & z^{-1}e^{-2\pi i/N} & z^{-2}e^{-4\pi i/N} & \dots & z^{-N+1}e^{-2\pi i(N-1)/N} \\ 1 & z^{-1}e^{-4\pi i/N} & z^{-2}e^{-8\pi i/N} & \dots & z^{-N+1}e^{-4\pi i(N-1)/N} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & z^{-1}e^{-2\pi i(N-1)/N} & z^{-2}e^{-4\pi i(N-1)/N} & \dots & z^{-N+1}e^{-2\pi i(N-1)^2/N} \end{bmatrix} \tag{13}$$

In equation 13, the z is a scale for the matrix that is not smaller than 1. If $z = 1$. Then the equation 13 becomes the matrix that calculates the Discrete Fourier Transform (DFT) for the trace signal. $\underline{F_z} d_{cal}(r,t,m)$ means that we are calculating the cross-correlation with different frequencies of sin and cos functions and resulting in the frequency spectrum. And the equation 12 means we are calculating the ℓ_2 norm frequency spectrum between the synthetic data and observed data as the objective function. If $z > 1$, then we are calculating the cross-correlation of the synthetic data with a series of different frequencies sin and cos function with damped amplitude. A larger value of z represents a larger damped property as time propagates. The correlation results would also ignore the later time steps of the shotrecords. Thus, the inversion strategy is that we truncate all the total inversion iterations into several segments, and in each segment, we use a different value of z, from large to small.

NUMERICAL TESTS

Objective function behaviors

In order to visualize the objective function behavior of the FWI, we plot how different objective functions react with the variation of the parameters of VP, VS, and density ρ. We perform forward modeling on a simple VP, VS, and density model, with homogeneous background and a small box shape object located in the center of the model. We would change the value of VP, VS, and ρ value in the small box and see how different kinds of objective functions react to these variations. We are going to test four objective functions, which are the ℓ_2 norm, ℓ_1 norm, the zero log correlation function, and the Multi-scale Z transform objective function.

Figure 1 shows the objective function behavior of the variation of parameter VP and VS. In the test, VP changes from $2500km/s \sim 4500km/s$ and VS changes from $500km/s \sim 2500km/s$. The valley in the center of Figure 1 shows the point for the true model, which has the smallest objective function value. We can clearly see that a lot of the local minimum existing in Figure 1, showing very strong non-linearity. However, the objective function of multi-scale Z transform is very smooth, showing no local minimum on the contour. In Figure 2, in the first row, we show the objective function behavior when we fix VP at $3300km/s$ and changes the VS. We can also see that the ℓ_2, ℓ_1, and correlation- objective

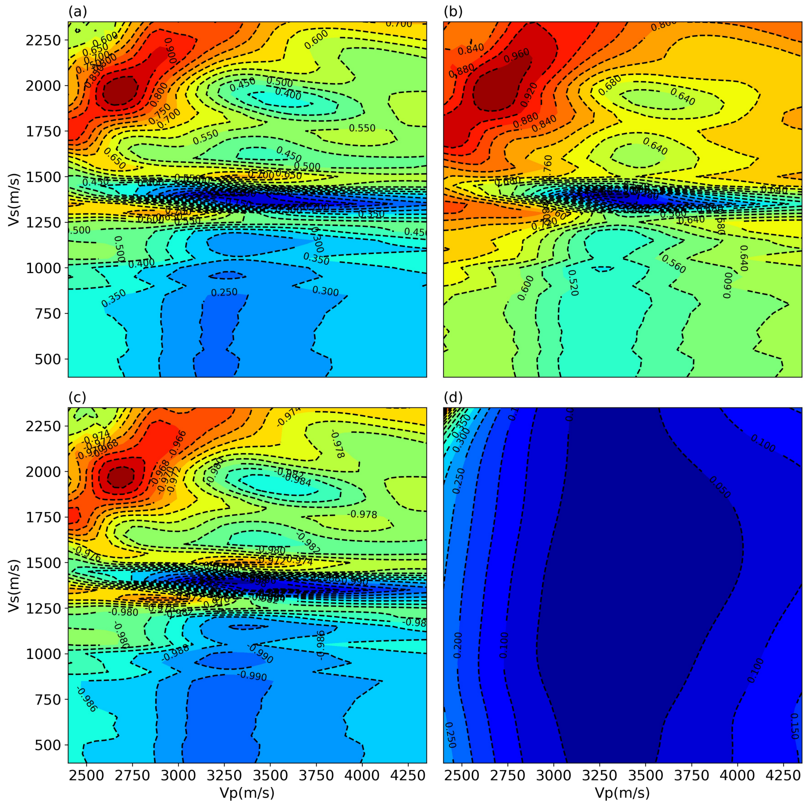

FIG. 1. Contour plot of the different objective functions with the variation of the VP and VS. (a) $\ell 2$ Norm . (b) $\ell 1$ Norm . (c) Zero lag correlation objective function. (d) Multi-scale Z transform

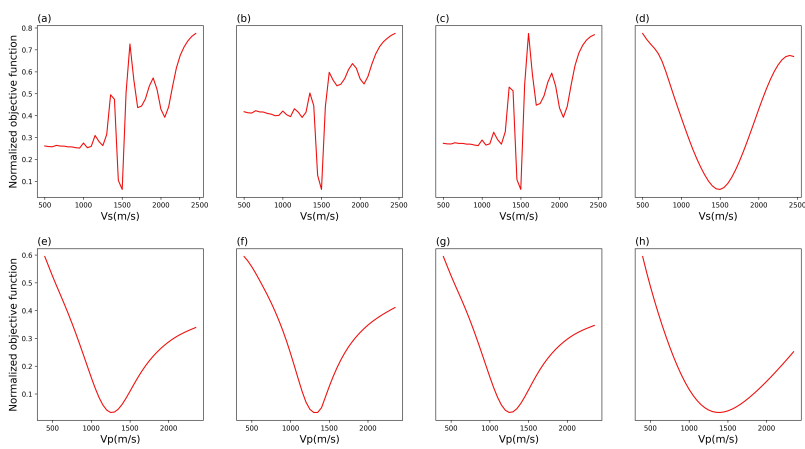

FIG. 2. The objective function responses for VP and VS variation. Figure (a)-(d) shows how $\ell 2$ norm, $\ell 1$ norm, GC and Multi-scale Z transform objective function changes with a fixed value of VP= $3300m/s$, and VS changes from $500m/s$ to $2250m/s$. Figure (e)-(h): the corresponding objective function changes with a fixed value of VS= $1300m/s$, and VP changes from $2500m/s$ to $4500m/s$.

functions have several local minimums, while the multi-scale Z transform has few. The second row in Figure 2 shows another story. If we fix VS and change VP, all the objective functions are very smooth and have no local minimum. It means that the VS has much more influence on the non-linearity of the objective function compared with VP.

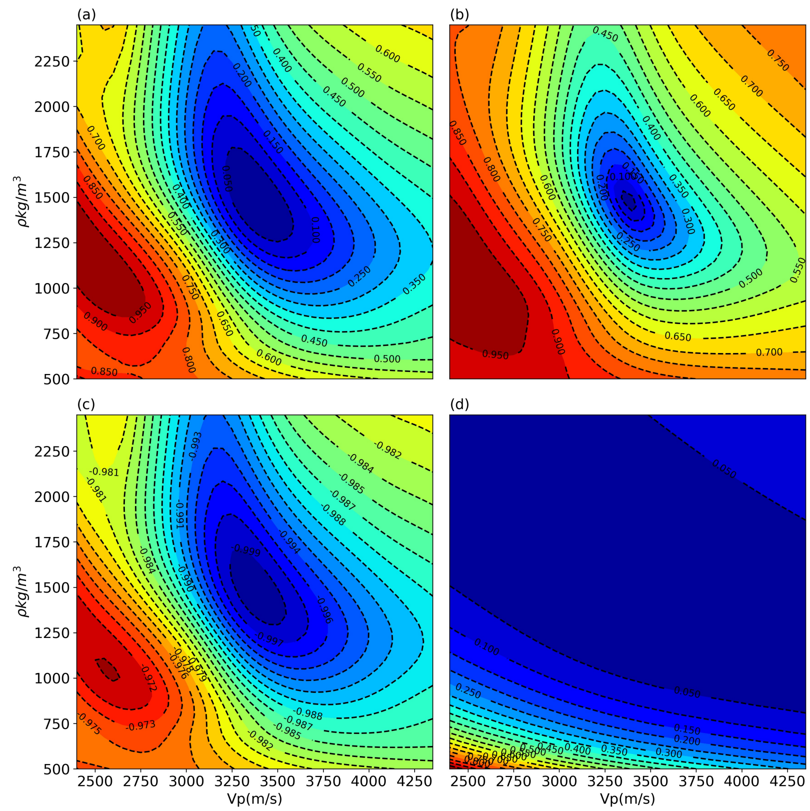

FIG. 3. Contour plot of the different objective functions with the variation of the VP and ρ. (a) $\ell 2$ Norm. (b) $\ell 1$ Norm. (c) Zero lag correlation objective function. (d) Multi-scale Z transform

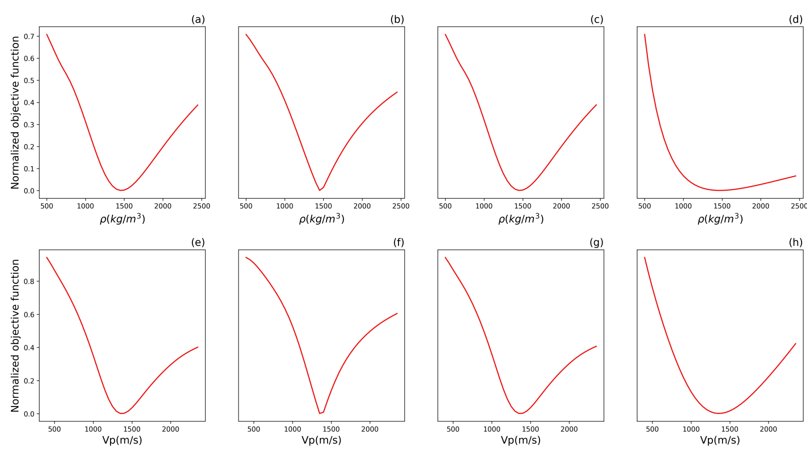

FIG. 4. The objective function responses for VP and ρ variation. Figure (a)-(d) shows how $\ell 2$ norm, $\ell 1$ norm, GC and Multi-scale Z transform objective function changes with a fixed value of Vp= $3300m/s$, and ρ changes from $500kg/m^3$ to $2500kg/m^3$. Figure (e)-(h): the corresponding objective function changes with a fixed value of $\rho = 500kg/m^3$, and VP changes from $2500m/s$ to $4500m/s$.

Figure 3 shows the contour plot of the objective function behavior for the variation of VP and ρ. The density ρ changes from $500km/m^3 \sim 2500km/m^3$ and VP still changes from $2500km/s \sim 4500km/s$. Still, the center of the contour shows true models, which have the smallest value. We can see that the objective functions show less local minimum compared with 1. However, if we look at the convergence for these objective functions, we can see that the objective function for multi-scale Z transform is less steep compared with the others. In, Figure 4 the first row shows that how objective function changes when we fix VP at $3300km/s$ and changes the density ρ. We observe that all the objective functions are very smooth. The same phenomenon could be seen when we change VP and fix density at $500kg/m^3$. And we could see that the convergence property of multi-scale Z transform objective function may not outperform the other objective functions. Without the influence for VS, the inversion is less likely to full into the local minimums.

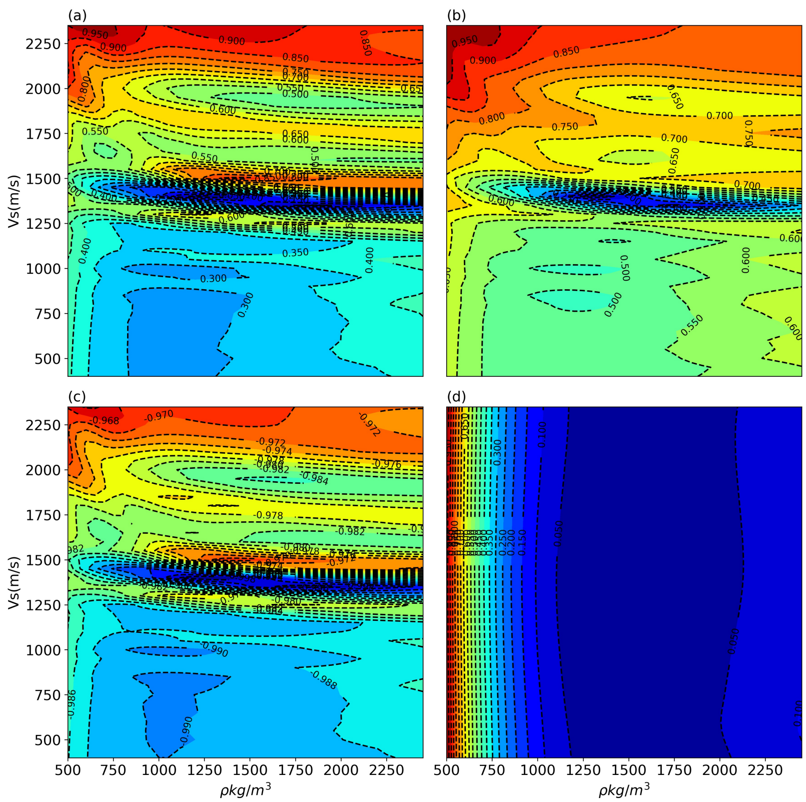

FIG. 5. Contour plot of the different objective functions with the variation of the VS and ρ. (a) $\ell2$ Norm . (b) $\ell1$ Norm . (c) Zero lag correlation objective function. (d) Multi-scale Z transform

Figure 5 shows how the objective function reacts to the variation of VS and density. Again, VS has a lot of influence on the objective functions. We can see a lot of local minimums on the contour, showing strong non-linearity. The "true valley" in the center of the image in the contour is very narrow, which can make inversion wonder out of the true convergence region. However, the contour given by multi-scale Z transform objective function shows a wider convergence region and showing no local minimum, which is helpful for the early state inversion for FWI. The first row in Figure 6 shows how the objective functions react to the variation of density when we fix VS and VP. It shows that there are no local minimums. However, if we change the value of VS and fix the VP and density, we can see a strong fluctuation of the objective function. Thus, we could also draw the conclusion that the variation of VS plays a much more important role in the non-linearity of the objective functions. An incorrect initial model or the incorrect update of the VS could severely influence the convergence property of the FWI.

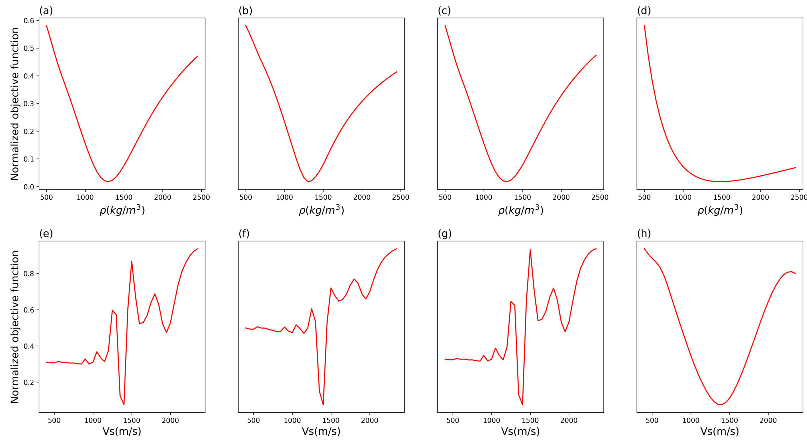

FIG. 6. The objective function responses for VS and ρ variation. Figure (a)-(d) shows how $\ell 2$ norm, $\ell 1$ norm, GC and Multi-scale Z transform objective function changes with a fixed value of $Vs = 1300m/s$, and ρ changes from $500kg/m^3$ to $2500kg/m^3$. Figure (e)-(h): the corresponding objective function changes with a fixed value of $\rho = 1500kg/m^3$, and VS changes from $500m/s$ to $2500m/s$.

Adjoint sources

By choosing the scale factors in equation 6, we can change decide the damping property of the sin and cos function that the data correlate with, which results in different kinds of adjoint sources in each sub iterations. For instance, in Figure 7, 8, and 9, shows the adjoint sources for the multi-scale Z transform objective function with different scale factor. We can see that with a larger scale factor in equation 13, the synthetic data would correlate with a function that has stronger damping with respect to time. As the surface waves always have rapid amplitude decaying properties, we can see in Figure 7 and Figure 8, that we can see surface-wave-like records are extracted. This means we are firstly using the surface waves to build the surface structure of the velocity models. Then, we will use reflection waves to perform the inversion, which is a multi-scale full waveform inversion. As the scale factor z gets bigger, we could see that more reflection-waves-like records could be observed in Figure 7, which means that reflections are going to be utilized in the inversion, which helps to build more details of the velocity models. When $z = 1$, the $\underline{F_z}$ is doing Fourier transform, which has no loss of information. We will use $z = 1$ in the last sub iteration for the inversion. 12 is this means we are using the full waveform to build the velocity model. Thus, the multi-scale Z transform objective functions have the ability to use different kinds of waveforms to achieve multi-scale full waveform inversion.

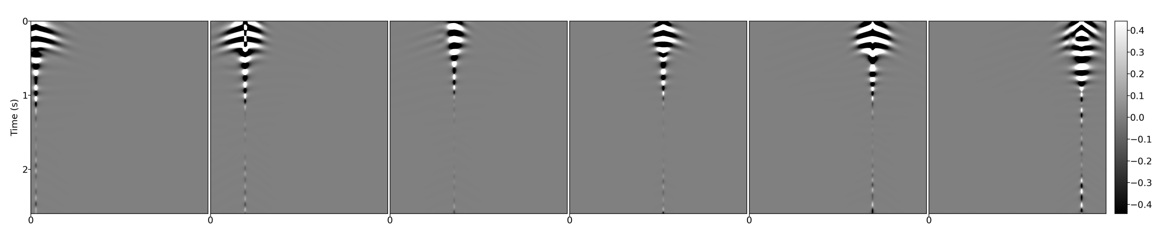

FIG. 7. Multi-scale Z transform adjoint sources scale=1.005

Figure 10 shows the adjoint source in the time domain. We can see that as the scale

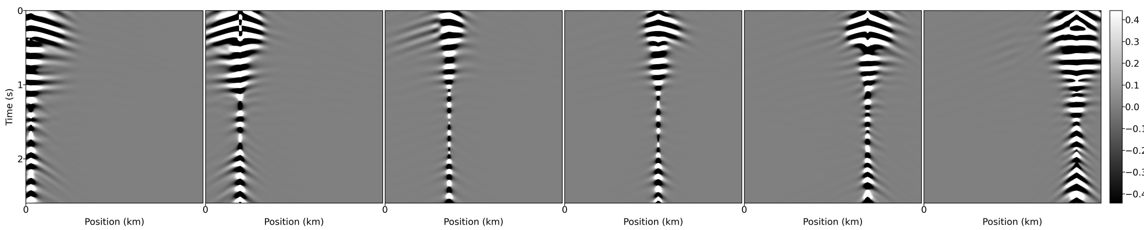

FIG. 8. Multi-scale Z transform adjoint sources scale=1.002

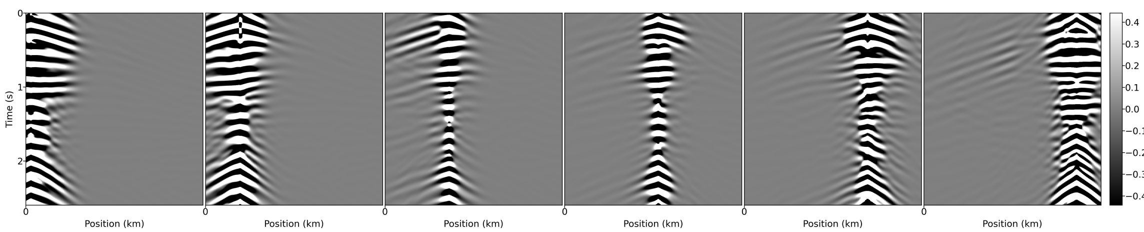

FIG. 9. Multi-scale Z transform adjoint sources scale=1.003

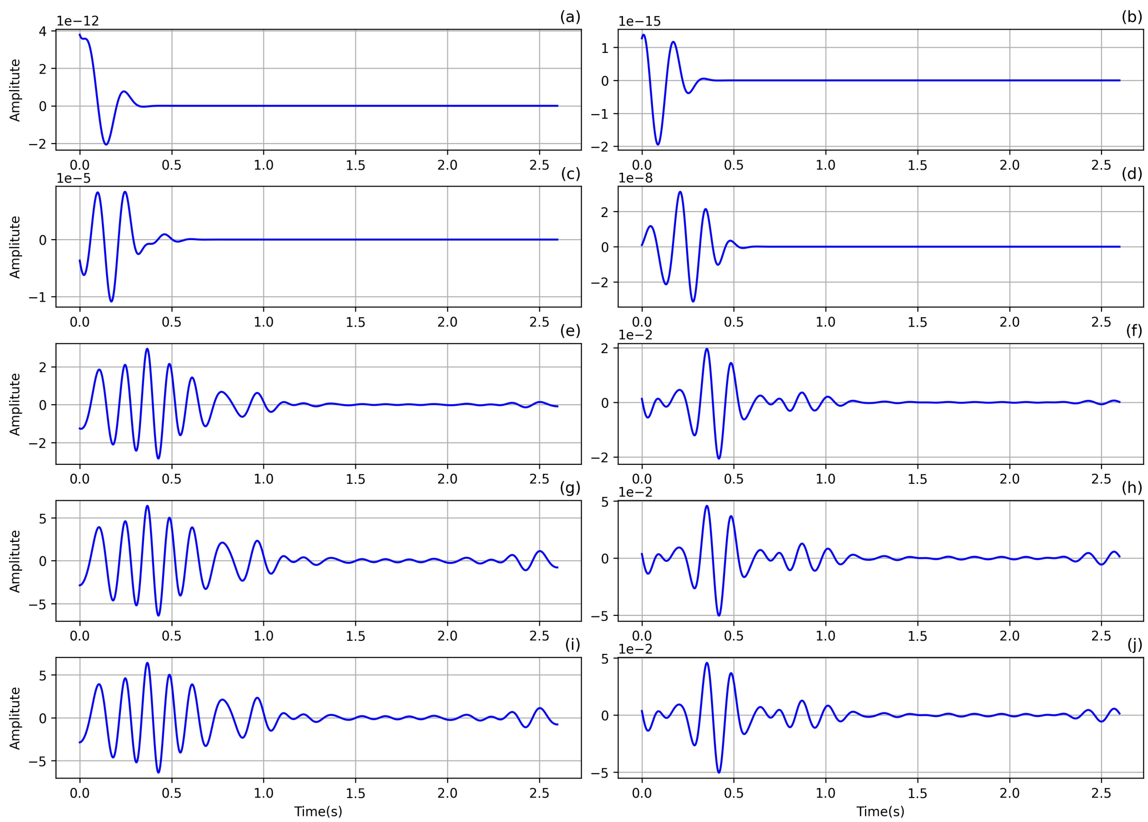

FIG. 10. Multi-scale Z transform adjoint sources with different scale values. (a)-(b) z and x adjoint sources with scale value=1.005. (c)-(d) z and x adjoint sources with scale value=1.001. (e)-(f) z and x adjoint sources with scale value=1.0003. (a)-(b) z and x adjoint sources with scale value=1.

factor changes from $z = 1.005$, to $z = 1.001$ and to $z = 1.0003$. More information could be observed. With $z = 1.0003$, the adjoint source of multi-scale Z transform contains almost

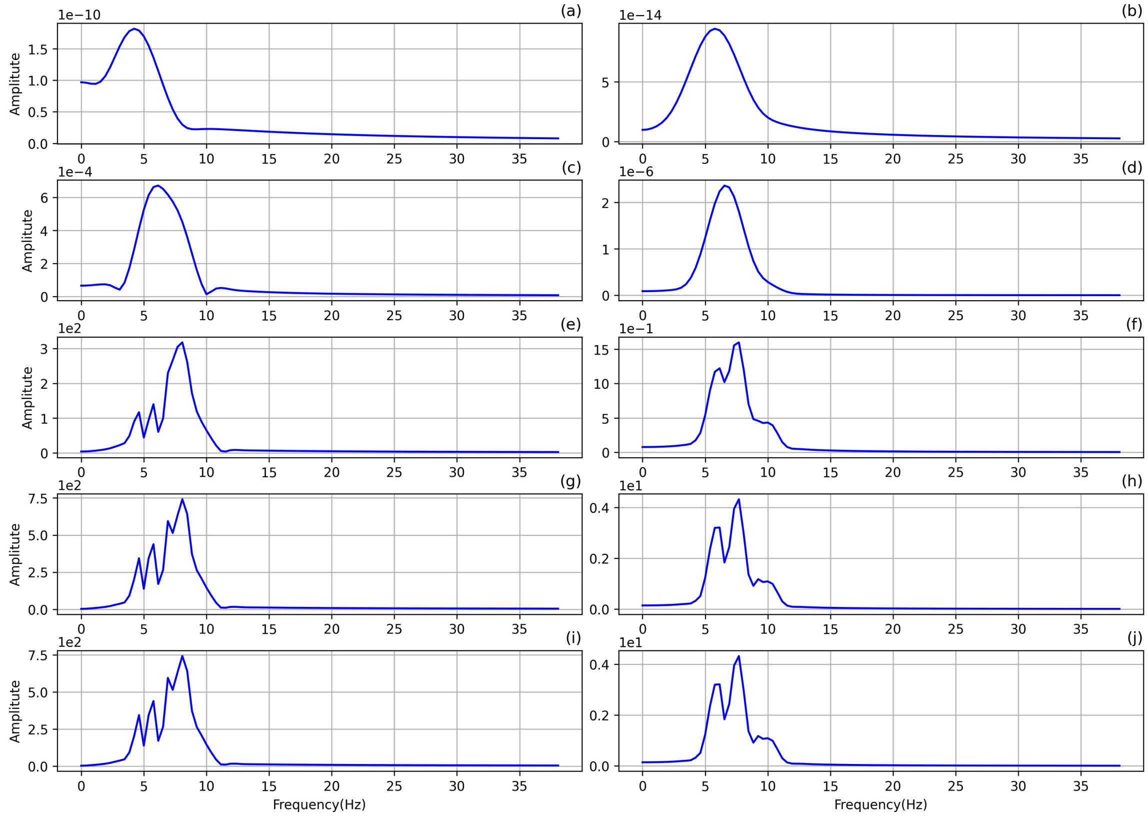

FIG. 11. Frequency spectrum in Figure 10. (a)-(b) z and x adjoint sources with scale value=1.005. (c)-(d) z and x adjoint sources with scale value=1.001. (e)-(f) z and x adjoint sources with scale value=1.0003. (a)-(b) z and x adjoint sources with scale value=1.

the same information with the ℓ_2 norm objective function. Figure 11 shows the frequency spectrum of the adjoint sources. We could clearly see that the adjoint source for multi-scale Z transform with $z = 1.005$, Figure 11 has abundant low frequency components. The main frequency in Figure 11 (a) is smaller than 5 Hz, while the in Figure 11 (c) it is around 6Hz. With the increase of the z in Figure 11 (e), the main frequency is around 8Hz. Thus, the multi-scale Z transform with different z values could gradually increase the main frequency of the adjoint sources, which could help the inversion to gradually build the velocity model from general to detail.

CONCLUSIONS

In this study, we first discuss the behavior of ℓ_2, ℓ_1, and the correlation-based objective functions. We demonstrate how the objective functions would react to the variation of the parameters for VP, VS, and density. We conclude that all the objective functions for ℓ_2, ℓ_1, and correlation-based objective functions contain several local minimums, and the convergence regions are all very small. The all these three objective functions are sensitive for variation of the VS parameter. When VS changes, the objective functions fluctuate a lot m indicating that VS is very hard to invert. We also introduce the multi-scale z transform objective function, which has a broader convergence region than the other three objective functions with few local minimums, which means that the inversion could perform well the initial model is not good. According to the frequency analysis of the adjoint sources, the multi-scale z transform contains fruitful information about the low frequencies. With the proper control of the scale factor z we could first use the surface waves to perform the inversion. With sufficient information about the low frequency in the records, we could build a reliable upper structure velocity model.

ACKNOWLEDGMENT

We thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19. The first author is also supported by the Chine Scholarship Council (CSC).

REFERENCES

Brossier, R., Operto, S., and Virieux, J., 2009, Seismic imaging of complex onshore structures by 2d elastic frequency-domain full-waveform inversion: Geophysics, **74**, No. 6, WCC105–WCC118.

Brossier, R., Operto, S., and Virieux, J., 2010, Which data residual norm for robust elastic frequency-domain full waveform inversion?: Geophysics, **75**, No. 3, R37–R46.

Bunks, C., Saleck, F. M., Zaleski, S., and Chavent, G., 1995, Multiscale seismic waveform inversion: Geophysics, **60**, No. 5, 1457–1473.

Choi, Y., and Alkhalifah, T., 2012, Application of multi-source waveform inversion to marine streamer data using the global correlation norm: Geophysical Prospecting, **60**, No. 4-Simultaneous Source Methods for Seismic Data, 748–758.

Choi, Y., and Alkhalifah, T., 2015, Unwrapped phase inversion with an exponential damping: Geophysics, **80**, No. 5, R251–R264.

Crase, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A., 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, No. 5, 527–538.

Djikpéssé, H. A., and Tarantola, A., 1999, Multiparameter l1 norm waveform fitting: Interpretation of gulf of mexico reflection seismograms: Geophysics, **64**, No. 4, 1023–1035.

Dong, S., Han, L., Zhang, P., and Mao, B., 2019, Waveform local travel-time corrected zero-mean cross-correlation-based fwi, *in* SEG Technical Program Expanded Abstracts 2019, Society of Exploration Geophysicists, 1551–1555.

Huber, P. J., 1973, Robust regression: asymptotics, conjectures and monte carlo: The annals of statistics, 799–821.

Liu, Y., Teng, J., Xu, T., Wang, Y., Liu, Q., and Badal, J., 2017, Robust time-domain full waveform inversion with normalized zero-lag cross-correlation objective function: Geophysical Journal International, **209**, No. 1, 106–122.

Pyun, S., Son, W., and Shin, C., 2009, Frequency-domain waveform inversion using an l 1-norm objective function: Exploration Geophysics, **40**, No. 2, 227–232.

Shah, N., Warner, M., Washbourne, J., Guasch, L., and Umpleby, A., 2012, A phase-unwrapped solution for overcoming a poor starting model in full-wavefield inversion, *in* 74th EAGE Conference and Exhibition incorporating EUROPEC 2012, European Association of Geoscientists & Engineers, cp–293.

Sirgue, L., and Pratt, R. G., 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, No. 1, 231–248.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM.

Wu, R.-S., Luo, J., and Wu, B., 2014, Seismic envelope inversion and modulation signal model: Geophysics, **79**, No. 3, WA13–WA24.