FAST LINE-SEARCH FOR FULL WAVEFORM INVERSION IN THE HUBER NORM

XIAONA MA

Institute of Geophysics, China Earthquake Administration, Beijing 100081, P.R. China.
kymxn0224@163.com

(Received July 8, 2019; revised version accepted February 10, 2021)

ABSTRACT

Ma, X., 2021. Fast line-search for full waveform inversion in the Huber norm. *Journal of Seismic Exploration*, 30: 347-364.

The misfit functions decide the robust performance of full-waveform inversion (FWI) when the data is contaminated by noise. Huber norm combines the L2 norm when residuals are small and L1 norm when residuals are large, which compromise the merits of them. It not only improves the anti-noise ability with the L1 norm, but also keeps smoothness for small residuals with the L2 norm. Line searches are always important for FWI process. Step-length can typically be calculated using an inexact or exact line-search method. Optimal step-length can prevent the over- or under-estimations, and make FWI to reach a global minimum along searching direction with fast convergence rate. Exact line searches for local optimization methods can be performed very efficiently for computing solutions in robust norms, thereby promoting convergence rate of FWI. Therefore, we derive an exact line search method, i.e., the analytical step-length method (ASLM), for the Huber norm. Through numerical tests on noise-free and noisy data of Overthrust model, we demonstrate the efficiency of ASLM for Huber norm. In addition, we also compare Huber with L2 norm on the data contaminated by non-Gaussian noise, such as ground-motion noise. We think that ASLM is an efficient optimal step-length estimation method for the Huber norm in FWI. Meanwhile, the Huber norm makes the FWI more robust than the L2 norm alone.

KEY WORDS: time-domain full waveform inversion, Huber norm,
exact line-search method, analytical step-length method.

INTRODUCTION

Full waveform inversion (FWI) is becoming a powerful geophysical tool to describe the subsurface geological structures, evaluate soil properties, monitoring CO_2 sequestration, and characterize energy reservoirs. Lailly (1983) and Tarantola (1984) originally proposed adjoint techniques to calculate the gradient of the data misfit, and many evolved approaches to

0963-0651/21/$5.00

waveform inversion algorithms have since appeared (Bunks et al., 1995; Pratt and Worthington, 1990; Pratt et al., 1998; Plessix, 2006; Shin and Cha, 2008). FWI has increasingly potential to achieve the high-resolution velocity estimate, but it suffers from difficulties such as non-linearity, ill-posedness, and high computational cost (Brossier et al., 2010); in addition, for the real seismic data we need face the obstacles including the absence of low-frequency data and noise (Ha et al., 2009; Liu et al., 2017).

FWI is a data-fitting procedure, aiming to minimize a misfit function that is defined as the difference between numerically calculated seismic data and observed data (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009). Noise is an important issue for FWI, because its present will affect the stable and convergence of the inversion procedure (Wang and Rao, 2006). And geophysicist have to face the reality that real seismic data is always polluted by many kinds of noise, such as ambient noise. The anti-noise ability of FWI mainly depends on the misfit functions. Researches think that the Huber norm can provide effective results in the presence of outliers and non-Gaussian noises (Guitton and Symes, 2003; Ha et al. 2009; Brossier et al., 2010). The Huber norm combines the L2 norm when residuals are small and L1 norm when residuals are large, thus compromising the merits of them. In this manner, it not only improves the anti-noise ability with L1 norm, but also keeps smoothness for small residuals with an L2 norm which can provide reliable results in the presence of Gaussian noise (Ha et al., 2009; Brossier et al., 2010).

The threshold in Huber norm controls the transition between these two different behaviours, which is an essential point providing the effectiveness of the practical application of the Huber norm for FWI. Bube and Langan (1997) proposed that the threshold value is to be roughly 0.6 times the standard deviation of the residual of the observed data and calculated data, or 0.8 times the mean deviation of residual. They thought that thus choosing is consistent with the assumptions made on the probability density by minimizing the misfit function. And Brossier et al. (2010) proposed that the threshold value is fixed to 0.2 times the mean of the observed data, and they thought that this setting is practically, and to be less sensitive to outliers in the data than the max value of the observed data indicated by Guitton and Symes (2003). This paper mainly study the step-length estimation method that is suitable for the Huber norm, so we do not provide the research about how to find the threshold. It remains an important issue for future study. Here we just adopt an empirical calculation method from previous works (Bube and Langan, 1997; Bube and Nemeth, 2007; Ha et al., 2009).

FWI generally begins with an adequate good initial model, then computes the search direction and step-length, updates the subsurface parameters iteratively, and eventually derives high-resolution models of the subsurface by exploiting the full information of wavefield records. Optimal step-length is necessary to reach a global minimum along search direction with fast convergence rate. Inexact or exact line-search techniques are currently used to compute the step-length for many meaningful algorithms for nonlinear minimization problems, especially nonlinear seismic inverse

problems (Pratt et al. 1998; Bube and Nemeth, 2007). To date, many researchers have studied the step-length estimation method. Backtracking line-search method (BLSM) is a typical inexact line-search method (Nocedal and Wright, 1999) and parabolic fitting search method through three points (PFSM-3) is a common exact line-search method (Operto et al., 2007; Vigh et al., 2009; Ma et al., 2019), they are both widely used in FWI (Hu et al., 2009; Dos Santos and Pestana, 2015; Operto et al., 2007; Tape et al., 2007; Vigh et al., 2009; Xu and McMechan, 2014; Liu et al., 2017). Considering the exact step-length estimation is closely related to the misfit function, in this paper, we propose the analytical step-length method (ASLM) for the Huber norm in FWI. Noise-free and noisy data simulated from Overthrust velocity model demonstrate that ASLM combined with Huber norm can obtain a high-resolution inverted velocity with lower computational cost compared with PFSM-3.

The paper is organized into four parts. The first section is the Introduction. In the next, the theory of FWI in the time domain is presented, and the ASLM of the Huber norm is derived. In the third section, noise-free and noisy data examples of ASLM and PFSM-3 in the Huber norm—are first presented to demonstrate the effectiveness of ASLM. Then, we compare the robustness of Huber-norm and L2-norm on the synthetic seismic data with strong ground-motion noise.

THEORY

Review of full waveform inversion in the time domain

The 2D constant density acoustic-wave equation in time domain is:

$$\left(\frac{\partial^2 p(\mathrm{x},\mathrm{z},t)}{\partial x^2}+\frac{\partial^2 p(\mathrm{x},\mathrm{z},t)}{\partial z^2}+s(\mathrm{x},\mathrm{z},t)\right)=\frac{1}{v^2}\frac{\partial^2 p(\mathrm{x},\mathrm{z},t)}{\partial t^2}, \qquad (1)$$

where t represents the time, (x,z) represents the position, $p(x,z,t)$ represents the pressure field, v represents the acoustic velocity, ∇^2 represents the Laplacian operator, and $s(x,z,t)$ represents a source term.

FWI is an optimization problem, aiming to minimize a misfit function f, which measures the difference between the observed data $p^{obs}(x,z,t)$ and calculated data $p^{cal}(x,z,t)$. We adopt the Huber norm as the misfit function (Huber, 1973):

$$f=\begin{cases}\dfrac{r^2}{2\varepsilon} & \text{for } 0\le|r|\le\varepsilon\\ |r|-\dfrac{\varepsilon}{2} & \text{for } \varepsilon<|r|\end{cases}, \qquad (2)$$

where ε is the threshold between the L1 norm and L2 norm, and

$$r = \sum_s \sum_g \left(p^{cal}(x,z,t) - p^{obs}(x,z,t) \right) \quad . \tag{3}$$

It is smooth near zero residual, weights small residuals by mean square, and treats large residuals with L1 norm (Fig. 1). Combining L1 and L2 norms in this manner, Huber norm becomes differentiable everywhere (Brossier et al., 2010).

The gradient of the Huber norm can be efficiently computed by the adjoint-state method (Plessix, 2006). Primarily, there are three steps in this process: forward propagation, backward propagation, and gradient calculation. The gradient formula can be expressed as:

$$g = \frac{\partial f}{\partial m} = \frac{2}{v(x,z)^3} \int_0^T \lambda \frac{\partial^2 p(x,z,t)}{\partial t^2} dt \,, \tag{4}$$

where λ represents the backward propagation wavefield, backward propagation residual can be expressed as:

$$backc(r) = \begin{cases} \dfrac{r}{\varepsilon} & \text{for } 0 \le |r| \le \varepsilon \\ sign(r) & \text{for } \varepsilon \le |r| \end{cases} \quad . \tag{5}$$

We find that the change of the misfit function only changes the adjoint source but not the entire gradient expression, so FWI has the similar gradient expression for the L2 and L1 norm.

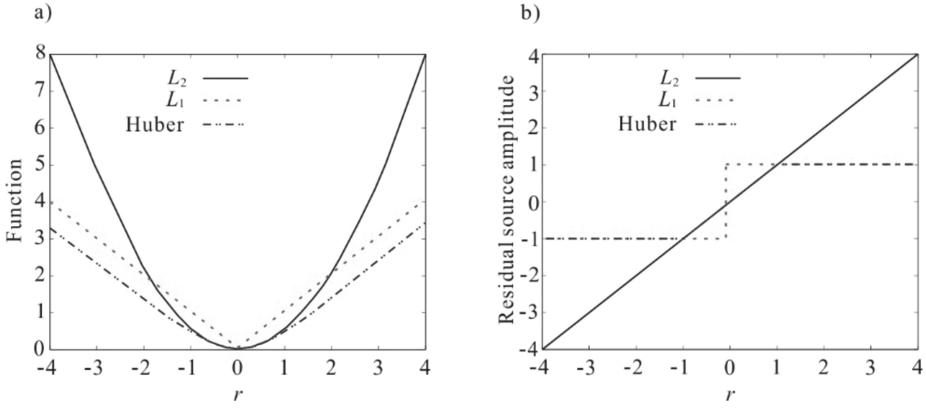


Fig. 1. a) The values of L1, L2, and Huber as functions; b) The associated residual source amplitude in the gradient expression (Brossier et al., 2010).

The velocity model can be updated by:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{d}_k \quad , \tag{6}$$

where $\mathbf{m}_{k+1}$ and $\mathbf{m}_k$ represent the (k+1)-th and k-th iteration model parameters, $\mathbf{d}_k$ represents the search directionand, α_k represents the step-length at the k-th iteration. The search direction is determined by the optimization methods, such as the steepest-descent method and the conjugate gradient, quasi-Newton methods, e.g., the limited memory BFGS (L-BFGS) method, truncated Newton solution or Newton method (Mora, 1987; Liu and Nocedal, 1989; Pratt et al., 1998; Brossier et al., 2009; Virieux and Operto, 2009; Warner et al., 2013; Hager and Zhang, 2006; Me´tivier et al., 2014). We use the L-BFGS optimization method to minimize the Huber misfit function because it can guarantee the convergence to global minima and entail less computational costs (Guitton and Symes, 2003).

Analytical step-length method of the Huber norm

If we adopt an optimal step-length estimation in FWI process, usually can improve the convergence rate and require fewer iterations to reach the global minimum. The exact step-length estimation is closely related to the misfit function. Therefore, we derive an optimal step-length formula for the Huber norm in FWI. In the current state, applying the Taylor-series expansion of α to the first order, the misfit function can be approximated as follows:

$$p^{cal}(c_k + \alpha_k d_k) \approx p^{cal}(c_k) + \alpha_k \left[\nabla p^{cal}(c_k)\right]^T d_k \quad . \tag{7}$$

So, we can obtain the misfit function in a new formula:

$$f \approx \begin{cases} \dfrac{\sum_s \sum_g \left(p^{cal}(c_k) + \alpha \left[\nabla p^{cal}(c_k)\right]^T d_k - p^{obs}(c_k) \right)^2}{2\varepsilon} & , \ 0 \le |r| \le \varepsilon \\ \sum_s \sum_g \sqrt{\left(p^{cal}(c_k) + \alpha \left[\nabla p^{cal}(c_k)\right]^T d_k - p^{obs}(c_k) \right)^2} - \dfrac{\varepsilon}{2} & , \ \varepsilon < |r| \end{cases} . \tag{8}$$

If the value of the misfit function approaches a local or global minimum, eq. (8) satisfies the following condition:

$$\frac{\partial f\left(c_k + \alpha_k d_k\right)}{\partial \alpha} = 0 \quad . \tag{9}$$

Thus, we can obtain the general form of the optimal step-length formula:

$$\alpha \approx \begin{cases} -\dfrac{\sum\sum\left[\nabla p^{cal}(c_k)\right]^T d_k \tilde{\delta p}}{\sum_s\sum_g\left[\nabla p^{cal}(c_k)\right]^T d_k\left[\nabla p^{cal}(c_k)\right]^T d_k}, & 0 \le |r| \le \varepsilon \\ -\alpha_t \dfrac{\sum_s\sum_g \tilde{\delta p}\,\tilde{\delta p}_t \Big/ \sqrt{\left(\tilde{\delta p}+c_1\left[\nabla p^{cal}(c_k)\right]^T d_k\right)^2}}{\sum_s\sum_g \tilde{\delta p}_t\,\tilde{\delta p}_t \Big/ \sqrt{\left(\tilde{\delta p}+c_1\left[\nabla p^{cal}(c_k)\right]^T d_k\right)^2}}, & \varepsilon < |r| \end{cases}, \quad (10)$$

where c_1 represents a coefficient, α_t represents the test step-length, which need satisfy the following condition:

$$\tilde{\delta p} = P^{cal}(x,z,t) - P^{obs}(x,z,t), \quad (11)$$

$$\tilde{\delta p}_t = P^{cal}(c_k + \alpha_t d_k) - P^{cal}(c_k), \quad (12)$$

and

$$\left[\nabla p^{cal}(c_k)\right]^T d_k \approx \frac{p^{cal}(c_k + \alpha_t d_k) - p^{cal}(c_k)}{\alpha_t}. \quad (13)$$

From eq. (10), we think that the calculation of ASLM need update the simulated data by a trial step-length to calculate the wavefield difference, and it must store the wavefield residual $\tilde{\delta p}$ of every shot at every iteration.

NUMERICAL EXAMPLES

In this section, noise-free and noisy synthetic data generated from Overthrust velocity model (Ravaut et al. 2004) are firstly adopted to demonstrate the efficiency and robustness of the ASLM with the Huber norm, then compare it with PFSM-3. Secondly, we compare the robustness of the Huber and L2 norms when non-uniform noise exists. Strong ground-motion noise is added to the synthetic data of the Overthrust model. In addition, for all numerical tests in this study, the threshold value ε for the Huber norm is fixed to be roughly 0.6 times the standard deviation of the data residual,

$$r = \sum_s\sum_g\left(p^{cal}(x,z,t) - p^{obs}(x,z,t)\right) \quad ,$$

thus choosing is indicated from previous work of Bube and Langan (1997).

To improve the chances of avoiding cycle skipping and to attain the global minimum rapidly for FWI in time domain, we adopt the multi-scale method to prevent the misfit function from falling into local minima (Bunks et al., 1995; Boonyasiriwat et al., 2010; Ma et al., 2017). The finite-difference method is employed in all examples for forward source-propagation and back-propagation of the residuals, and we apply the perfectly-matched-layer (PML) condition at boundaries to absorb the artificial interference (Bérenger, 1994).

Noise free data

Here, we mainly demonstrate the efficiency of ASLM, and compare it with PFSM-3 for the Huber norm by the tests on noise-free data generated from Overthrust velocity model. FWI is highly nonlinear and the results strongly depend on the accuracy of the initial model in the framework of local optimization (Brossier et al., 2010). So we use the smooth velocity model and linearly increasing velocity model, respectively, as the initial model to test the robustness of this step-length estimation method in FWI process.

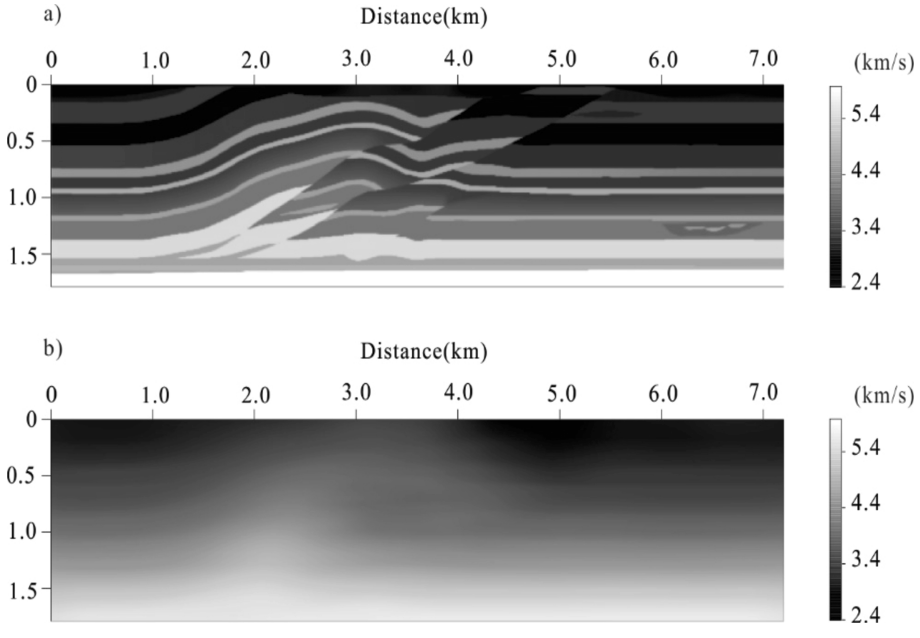


Fig. 2. a) True Overthrust velocity model; b) Smooth velocity model used as initial model.

The Overthrust velocity model used in all examples is resized to 601×150 grid samples in the horizontal and vertical directions while the grid spacing is 12 m in each direction. The synthetic data are generated from 60 shots separated by an interval of 120 m. To record each shot, 601 receivers are regularly distributed in the horizontal direction at 12 m intervals. The source signal is a Ricker wavelet with a central frequency of 18 Hz. The sampling interval is 1 ms, and the total recording length is 3 s. The acquisition geometry of all numerical tests is same. Fig. 2a shows the Overthrust velocity model used to synthesize the observed data, Fig. 2b shows the smooth velocity model used as the initial model.

Figs. 3a-3b show the inversion results using the ASLM and PFSM-3. Compared with the true Overthrust velocity model in Fig. 2a, we find that ASLM recover all the features and structures of the Overthrust velocity appropriately, which is comparable with the inverted velocity model obtained using PFSM-3. To see how much information our algorithm recovered, we show velocity profiles extracted at the horizontal position of 3.0 km, 4.8 km, and 6.0 km, respectively, in Figs. 4a-4c. The inverted velocity profiles of ASLM and PFSM-3 appear to show a trend similar to the true velocity profile. Obviously, from the analysis of inversion results, we think that the inversion accuracy of ASLM is satisfied.

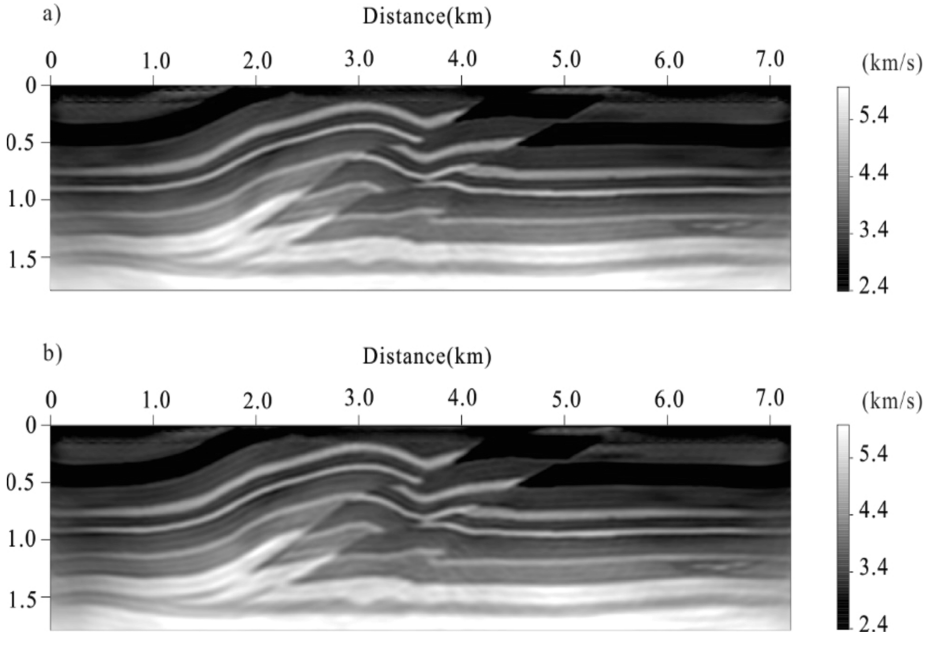


Fig. 3. Inverted velocity models obtained by using a) ASLM and b) PFSM-3 in the Huber-norm. Smooth velocity model in Figure2b is used as initial model.

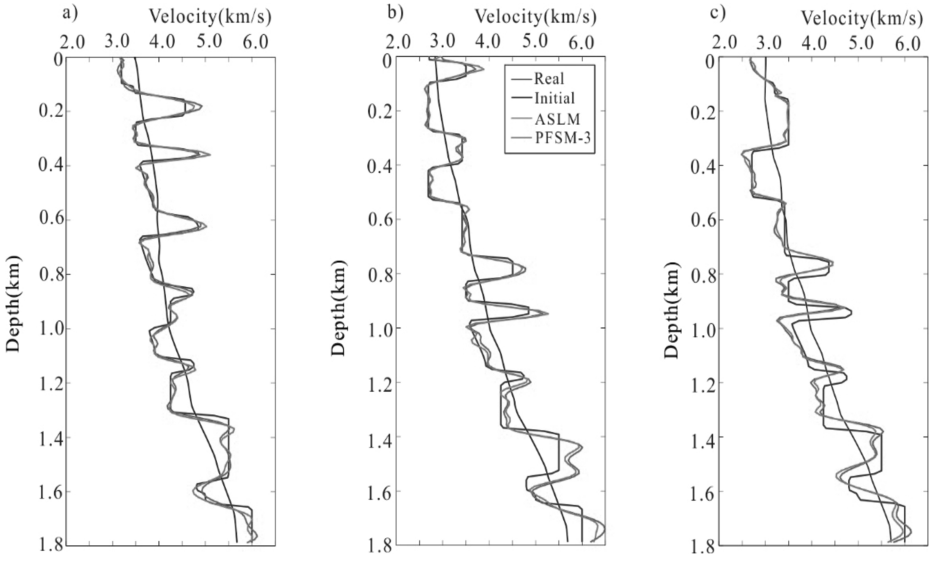


Fig. 4. Velocity profiles at horizontal positions of a) 3.0, b) 4.8 and c) 6.0 km. Blue represents the true model (Fig. 2a), black is the initial model (Fig. 2b), green is the model inverted using the ASLM(Figure 3a), and red is the model inverted using the PFSM-3 (Fig. 3b).

To compare the convergence rates, we calculate the normalized root-mean-square (rms) error of the velocity model through the inversion (Fig. 5a). From this figure, we can see that ASLM obtain satisfied FWI result and converge faster than PFSM-3. While Fig. 5b presents the iteration number and the number of step-length attempts (each attempt means an extra forward modeling) required by the two step-length estimation methods. The iteration numbers of ASLM are 79 which are smaller than those of PFSM-3. And PFSM-3 need at least two extra misfit function calculations to get the optimal step-length at every iteration. ASLM also need one extra attempt to calculate the optimal step length. Therefore, the extra forward modeling calculations of PFSM-3 and ASLM are 79 and 195, respectively. From the above analysis, it is found that the total forward modeling costs of the ASLM is less.

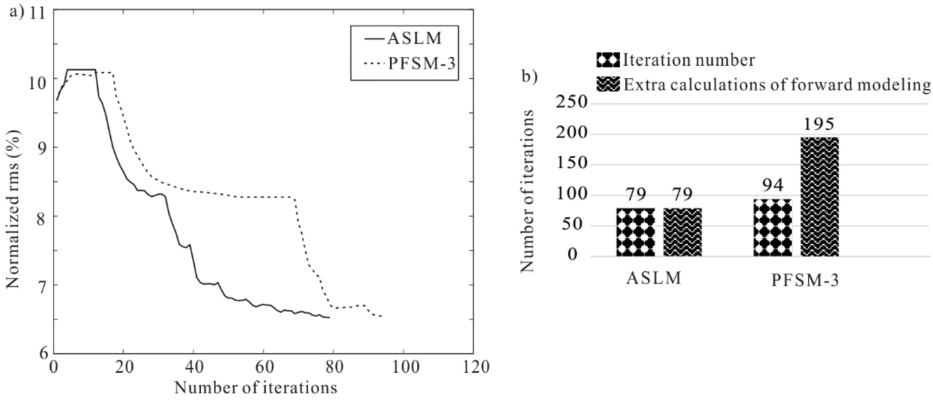


Fig. 5. a) Normalized rms errors of the FWI inverted model by using ASLM and PFSM-3; b) total number of iterations and number of extra calculations of forward modeling to achieve the final results by using ASLM and PFSM-3.

To test our method with different initial approximations for Overthrust model in FWI, here we choose the linearly increasing velocity model as the initial model (Fig. 6), which delete some macro velocity features compared with the smooth initial velocity model in Fig. 2b. The velocity model obtained from ASLM and PFSM-3 is shown in Fig. 7. We can find that the inversion result of ASLM is comparable with that of PFSM-3. This expectation is confirmed by the normalized rms error of the velocity model through the inversion in Fig. 8. Although both methods are successful in recovering fine details of the true velocity model, ASLM proved more efficient, due to the iteration number and the number of failed step-length attempts are smaller than PFSM-3.

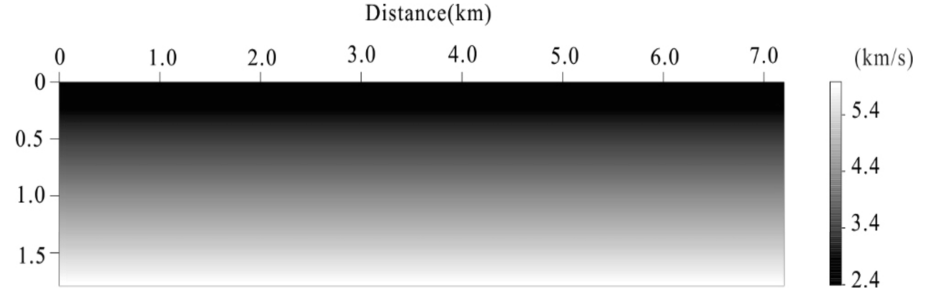


Fig. 6. Linearly increasing velocity model used as the initial model.

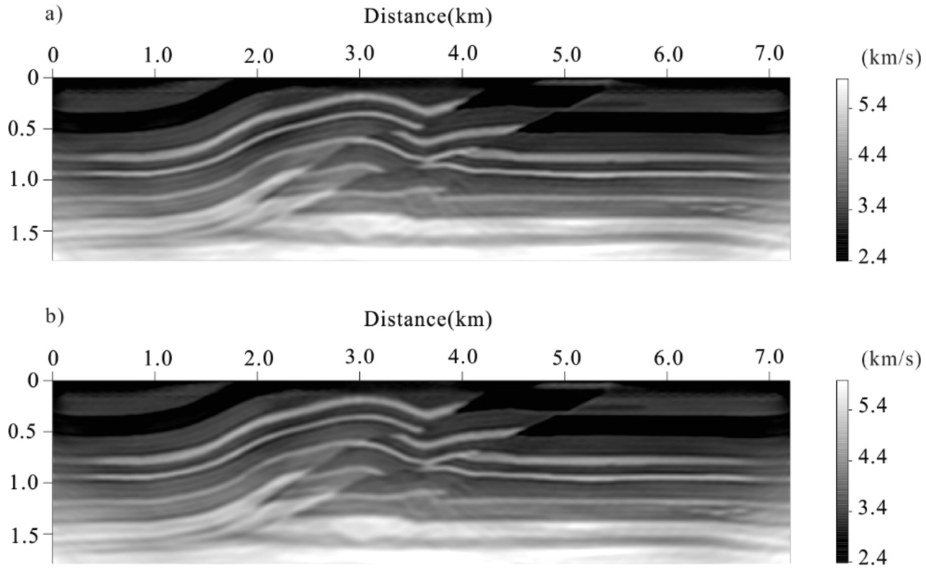


Fig. 7. Inverted velocity models obtained by using a) ASLM and b) PFSM-3 in the Huber-norm. Linearly increasing velocity in Fig. 6 is used as the initial model.

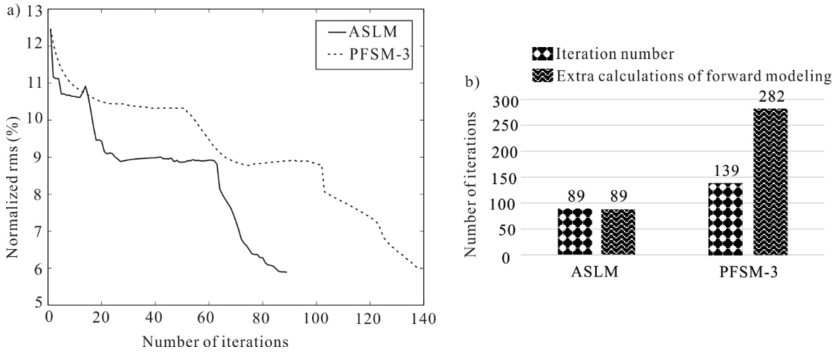
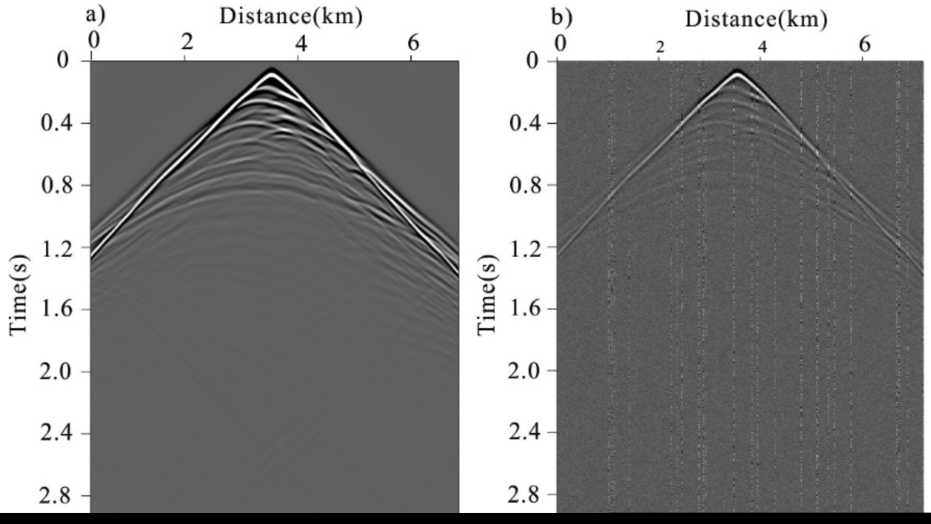


Fig. 8. a) Normalized rms errors of the FWI inverted model by using ASLM and PFSM-3 in Fig. 7; b) total number of iterations and number of extra calculations of forward modeling to achieve the final result by using ASLM and PFSM-3.

Considering that FWI is sensitive to the accuracy of the initial model in the framework of local optimization, we use two different initial velocity models, smooth velocity model and linearly increasing velocity model, to test the efficiency of ASLM. The above noise-free data examples both demonstrate the robustness of ASLM we proposed in this paper. Although ASLM and PFAM-3 both are successful in recovering fine details of the true velocity model, ASLM proves more computational efficient.

Noisy data test

We now test the proposed method on a data set contaminated with strong ground-motion noise, generating synthetic data from Overthrust velocity model in Fig. 2a. We make use of pseudorandom number to choose some seismic traces rescaling by a factor of 15, as well as contaminated by Gaussian white noise (S/N = 20 dB), thus simulating a poorly pre-processed strong ground-motion dataset (Liu et al., 2017). Noise-free data set and data contaminated with strong ground-motion noise are displayed in Fig. 9.



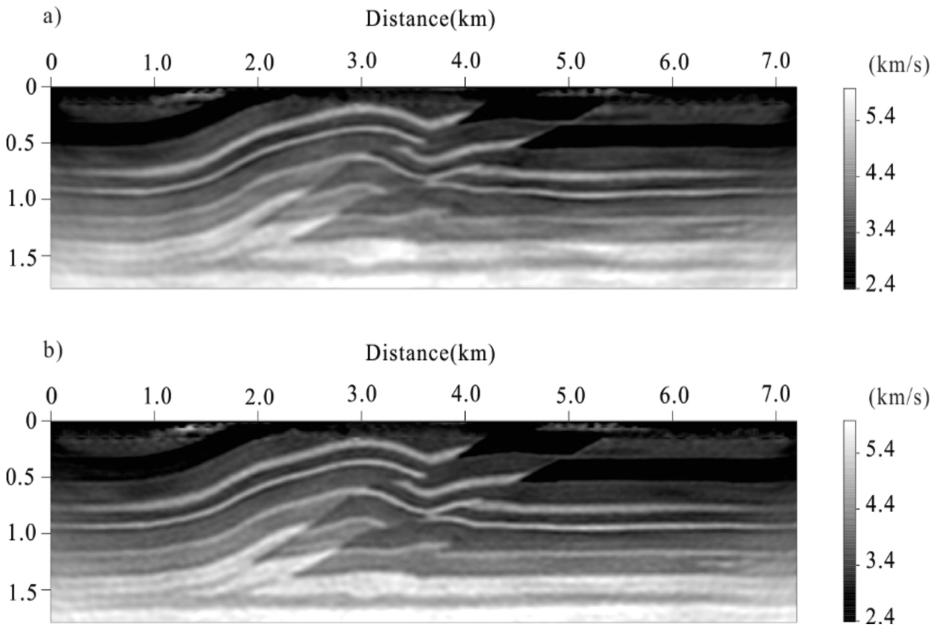


Fig. 10. Inverted velocity models obtained from data contaminated with strong ground-motion noise by using a) ASLM and b) PFSM-3 in the Huber-norm. Smooth velocity model in Fig. 2b is used as initial model.

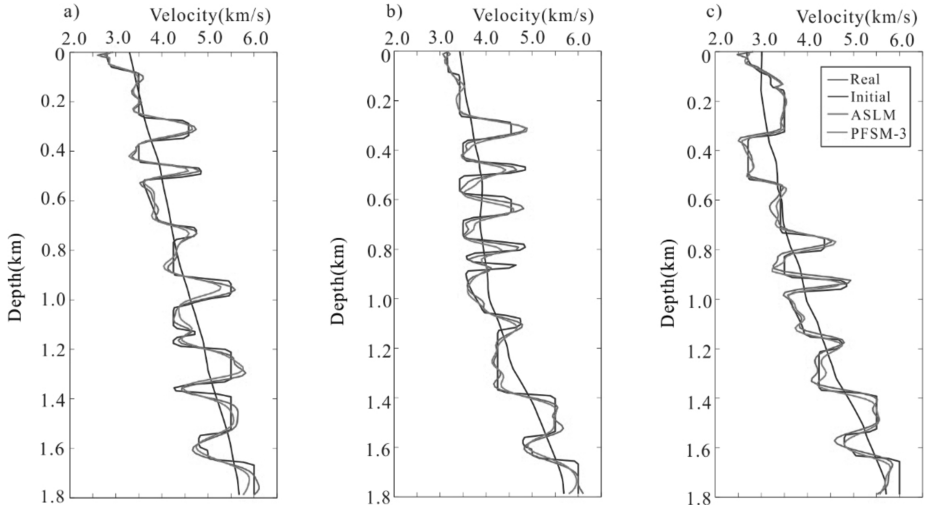


Fig. 11. Velocity profiles at horizontal positions of a) 3.0, b) 4.8 and c) 6.0 km. Blue represents the true model (Fig. 2a), black is the initial model (Fig. 2b), green is the model inverted using the ASLM (Fig. 10a), and red is the model inverted using the PFSM-3 (Fig. 10b).

Fig. 12a shows the normalized rms error of the velocity model through the inversion. For the noisy data set, ASLM converge faster than PFSM-3, but PFSM-3 provides slightly better FWI result. Fig. 12b presents the iteration number and the number of extra attempts required by the two methods, it is found that the total forward modeling costs of the ASLM method is lower.

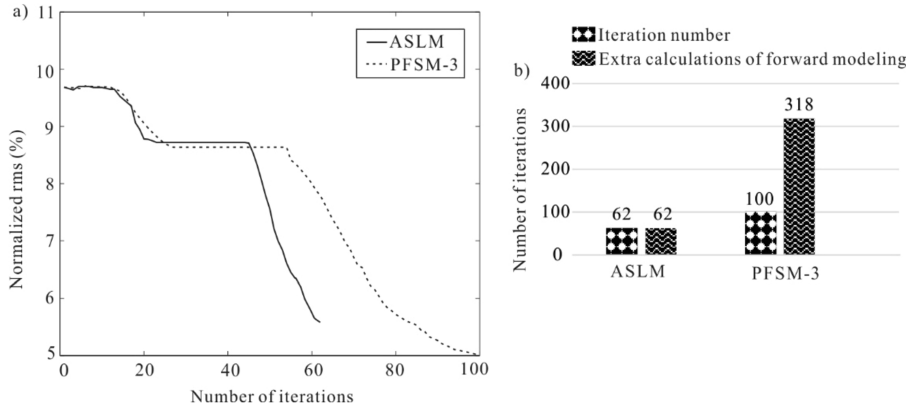


Fig. 12. a) Normalized rms errors of the FWI inverted model by using ASLM and PFSM-3 in Fig. 10; b) total number of iterations and number of extra calculations of forward modeling to achieve the final result by using ASLM and PFSM-3.

The data set contaminated with strong ground-motion noise in this FWI process again demonstrate the robustness of ASLM. And we think that in terms of the computational cost, ASLM performs outstanding in Huber norm.

Comparisons of Huber and L2

Here we compare the efficiency between the Huber and L2 norms. For the Huber norm, we use the ASLM because of the satisfactory results obtained in the above examples, and compared the results obtained with the L2 norm combined with the ASLM under the same circumstance (Ma et al., 2019). Ground-motion noise dataset is adopted to compare the Huber and L2 norms (Fig. 9b). Here we use the smooth velocity model in Fig. 2c as the initial model to implement this FWI process. The inversion results of the two norms are shown in Fig. 13. We can find that the Huber and L2-norm can both recover the Overthrust velocity model, but the structures and features of the Huber norm are reconstructed more properly. We show the velocity profiles of the true, initial, and inverted velocities in Fig. 14. From this we can see that the inversion result of L2 norm matches worse (indicated by arrows) than Huber norm. Fig. 15 shows the normalized rms error of the velocity model through the inversion. From this figure, we can see that Huber norm provide better FWI result even though it needs more iteration number.

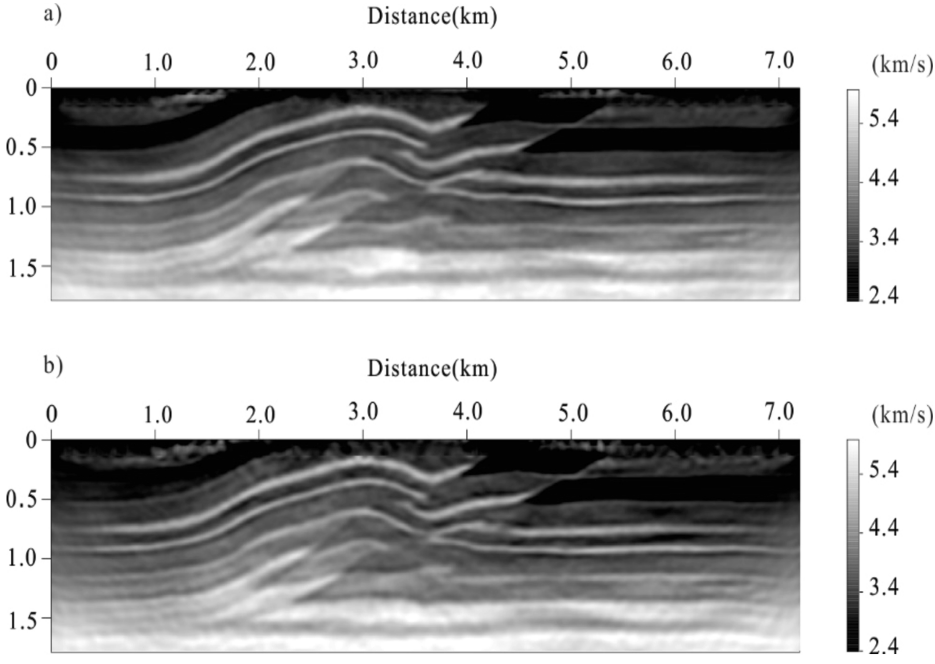


Fig. 13. Inverted velocity models obtained from data contaminated strong ground-motion noise by using a) Huber norm and b) L2 norm. Smooth velocity model in Fig. 2b is used as initial model.

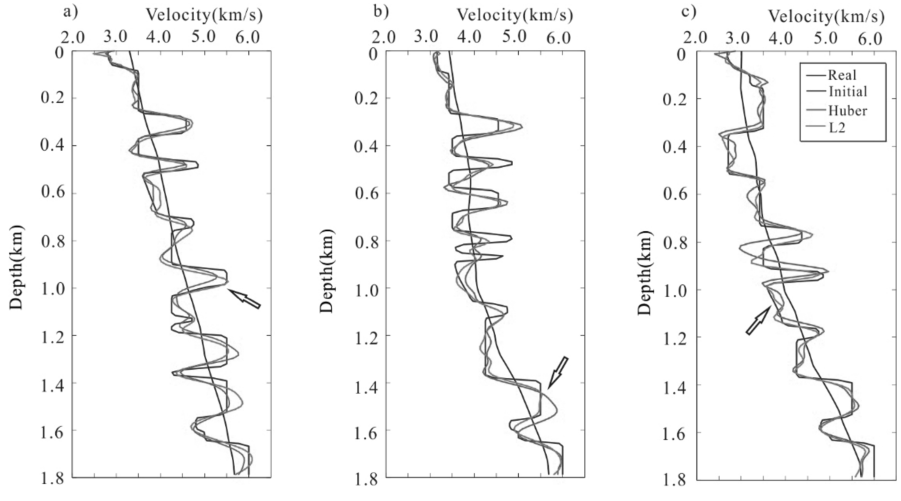


Fig.14 Velocity profiles at horizontal positions of a) 2.4, b) 3.6 and c) 6.0 km. Blue represents the true model (Fig. 2a), black is the initial model (Fig. 2b), green is the model inverted using the Huber norm (Fig. 13a), and red is the model inverted using the L2 norm (Fig. 13b).

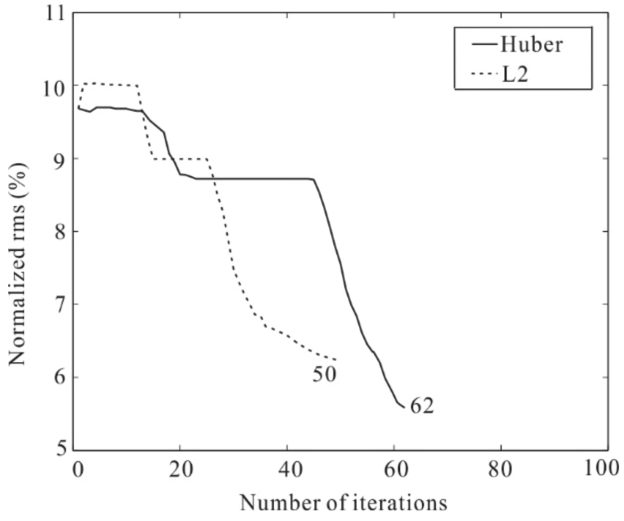


Fig. 15. Normalized rms errors of the FWI inverted model using Huber and L2.

As expected, the above examples presented here further indicate that the Huber norm effectively suppresses non-Gaussian noise in the inversion process. Huber norm can provide more reliable results than the L2 norm that can provide reliable results in the presence of Gaussian noise.

CONCLUSIONS

FWI starts from an initial model, then calculates the search direction and step-length, and updates the target velocity model iteratively by minimizing the misfit function. In this inversion process, step-length affects the convergence quality and computational cost, misfit function decides the tolerance of noise. Apparently, the two key factors both have important influences for the successful inversion.

Exact line-search method - ASLM, we propose in this paper for the Huber norm, can improve the computational efficiency while maintain the inversion accuracy. From the noise-free data examples, we demonstrate that ASLM performs better than PFSM-3 in Huber norm in terms of computational cost. The proposed method is also tested on synthetic data contaminated with outlier noise. ASLM can obtain a high-resolution result with less extra forward modeling than PFSM-3 in the presence of non-Gaussian noise.

Huber norm combines the L2 norm when residuals are small and L1 norm when residuals are large, thus compromising the anti-noise ability of L1 norm and the stability of L2 norm. When testing the sensitivity of FWI

against data in the presence of non-Gaussian noise, the inversion results obtained with the Huber norm are satisfactory, the robust Huber norm can give rise to high-resolution images. However, FWI with the classic L2 norm fails to invert the acceptable velocity models when the data contaminated by non-Gaussian noise.

ACKNOWLEDGEMENTS

This research is financially supported by the Special Fund of the Institute of Geophysics, China Earthquake Administration (Grant Number DQJB21K31). We thank the associated editors and anonymous reviewers for their careful reviewing and constructive suggestions.

REFERENCES

Bérenger, J.P., 1994. A perfectly matched layer for the absorption of electromagnetic waves. J. Computat. Phys., 114: 185-200.

Bunks, C., Saleck, F.M., Zaleski, S. and Chavent, G., 1995. Multiscale seismic waveform inversion. Geophysics, 60: 1457-1473.

Bube, K.P. and Nemeth, T., 2007. Fast line searches for the robust solution of linear systems in the hybrid and Huber norms. Geophysics, 72(2): A13-A17.

Boonyasiriwat, C., Valasek, P., Routh, P., Cao, W., Schuster, G.T. and Macy, B., 2009. An efficient multiscale method for time-domain waveform tomography. Geophysics, 74(6): WCC59-WCC68.

Brossier, R., Operto, S. and Virieux, J., 2009. Robust elastic frequency domain full waveform inversion using the L1 norm. Geophys. Res. Lett., 36: L20310.

Brossier, R., Operto, S. and Virieux, J., 2010. Which data residual norm for robust elastic frequency-domain full waveform inversion? Geophysics, 75(3): R37-R46.

Crase, E., Pica, A., Noble, M., McDonald, J. and Tarantola, A., 1990. Robust elastic nonlinear waveform inversion: Application to real data. Geophysics, 55: 527-538.

Djikpéssé, H.A. and Tarantola, A., 1999. Multiparameter L1 norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms. Geophysics, 64: 1023-1035.

Dos Santos, A.W.G. and Pestana. R.C., 2015. Time-domain multiscale full-waveform inversion using the rapid expansion method and efficient step-length estimation. Geophysics, 80(4): R203-R216.

Guitton, A. and Symes, W.W., 2003. Robust inversion of seismic data using the Huber norm. Geophysics, 68: 1310-1319.

Huber, P.J., 1973. Robust regression: asymptotics, conjectures, and Monte Carlo. Ann. Statist., 1: 799-821.

Hager, W.W. and Zhang, H.C., 2006. A survey of nonlinear conjugate gradient methods. Pacif. J. Optimiz., 2: 35-58.

Ha, T., Chung, W. and Shin, C., 2009. Waveform inversion using a backpropagation algorithm and a Huber function norm. Geophysics ,74(3): R15-R24.

Liu, D. and Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. Mathemat. Programm., 45: 503-528.

Liu, Y.S., Teng, J.W., Xu, T., Wang, Y.H., Liu, Q.Y. and Jose, B., 2017. Robust time-domain full waveform inversion with normalized zero-lag cross-correlation objective function. Geophys. J. Internat., 209: 106-122.

Liu, Y.S., Teng, J.W., Xu, T., Jose, B., Liu, Q.Y. and Zhou, B., 2017. Effects of conjugate gradient methods and step-length formulas on the multiscale full waveform inversion in time domain: Numerical Experiments. Pure Appl. Geophys., 174: 1983-2206.

Lailly, P., 1983. The seismic inverse problems as a sequence of beforenstack migration, Proc. Conf. Inverse Scattering, Theory and Application. Soc. Industr. Appl. Mathemat.: 206-220,

Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data. Geophysics, 52: 1211-1228.

Me´tivier, L., Bretaudeau, F., Brossier, R., Virieux, J. and Operto, S., 2014. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures. Geophys. Prosp. 62: 1353-1375.

Ma, X.N., Li, Z.Y., Xu, S.H., Ke, P. and Liang, G.H., 2017. Comparison of frequency-band selection strategies for 2D time-domain acoustic waveform inversion. J. Seismic Explor., 26: 499-519.

Ma, X.N., Li, Z.Y., Ke, P., Xu, S.H., Liang, G.H. and Wu, X.Q., 2019. Research of step-length estimation methods for full waveform inversion in time domain. Explor. Geophys., 50: 583-599.

Nocedal, J. and Wright, S.J., 1999. Numerical Optimization. Springer Verlag, Berlin.

Pica, A., Diet, J. and Tarantola, A., 1990. Nonlinear inversion of seismic reflection data in a laterally invariant medium. Geophysics, 55: 284-292.

Pratt, R.G. and Worthington, M.H., 1990. Inverse theory applied to multisource cross-hole tomography. Part I: Acoustic wave-equation method. Geophys. Prosp., 38: 287-310.

Pratt, R.G., Shin, C. and Hicks, G.J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. Geophys. J. Internat., 133: 341-362.

Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. Geophys. J. Internat., 167: 495-503.

Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A. and Dell' Aversana, P., 2004. Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt. Geophys. J. Internat., 159: 1032-1056.

Ren, Z.M., Liu, Y. and Zhang, Q.S., 2014. Multiscale viscoacoustic waveform inversion with the second generation wavelet transform and adaptive time–space domain finite-difference method. Geophys. J. Internat., 197: 948-974.

Shin, C. and Cha, Y.H., 2008. Waveform inversion in the Laplace domain. Geophys. J. Internat., 173: 922-931.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation. Geophysics, 49: 1259-1266.

Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data. Geophysics, 51: 1893-1903.

Tape, C., Liu, Q.Y. and Tromp, J., 2007. Finite-frequency tomography using adjoint methods - Methodology and examples using membrane surface waves. Geophys. J. Internat., 168: 1105-1129.

Hu, W.Y., Abubakar, A. and Habashy, T.M., 2009. Simultaneous multifrequency inversion of full-waveform seismic data. Geophysics, 74(2): R1-R14.

Vigh, D., Starr, E.W. and Kapoor, J., 2009. Developing earth models with full waveform inversion. The Leading Edge, 28: 432-435.

Virieux, J. and Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics. Geophysics, 74(6): WCC1-WCC26.

Wang, Y. and Rao, Y., 2006. Crosshole seismic waveform tomography I: Strategy for real data application. Geophys. J. Internat., 166: 1224-1236.

Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., Tekl, I., Guasch, L., Win, C., Conroy, G. and Bertrand, A., 2013. Anisotropic 3D full-waveform inversion. Geophysics, 78(2): R59-R80.

Xu, K. and McMechan, G.A., 2014. 2D frequency-domain elastic full-waveform inversion using time-domain modeling and a multistep-length gradient approach. Geophysics, 79(2): R41-R53.