



OPEN ACCESS

EDITED BY

Yong Zheng,
China University of Geosciences
Wuhan, China

REVIEWED BY

Jing Ba,
Hohai University, China
Kai Gao,
Los Alamos National Laboratory (DOE),
United States

*CORRESPONDENCE

Bing-Shou He,
hebinshou@ouc.edu.cn

SPECIALTY SECTION

This article was submitted to
Solid Earth Geophysics,
a section of the journal
Frontiers in Earth Science

RECEIVED 18 September 2022

ACCEPTED 23 November 2022

PUBLISHED 27 January 2023

CITATION

Dai M-X, He B-S and Huang W-S (2023),
Studies on modified limited-memory
BFGS method in full
waveform inversion.
Front. Earth Sci. 10:1047342.
doi: 10.3389/feart.2022.1047342

COPYRIGHT

© 2023 Dai, He and Huang. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Studies on modified limited-memory BFGS method in full waveform inversion

Meng-Xue Dai¹, Bing-Shou He^{2,3*} and Wei-Sen Huang¹

¹Engineering Research Center of Nuclear Technology Application, Ministry of Education, East China University of Technology, Nanchang, China, ²Evaluation and Detection Technology Laboratory of Marine Mineral Resources, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, ³Key Lab of Submarine Geoscience and Prospecting Techniques, Ministry of Education, Ocean University of China, Qingdao, China

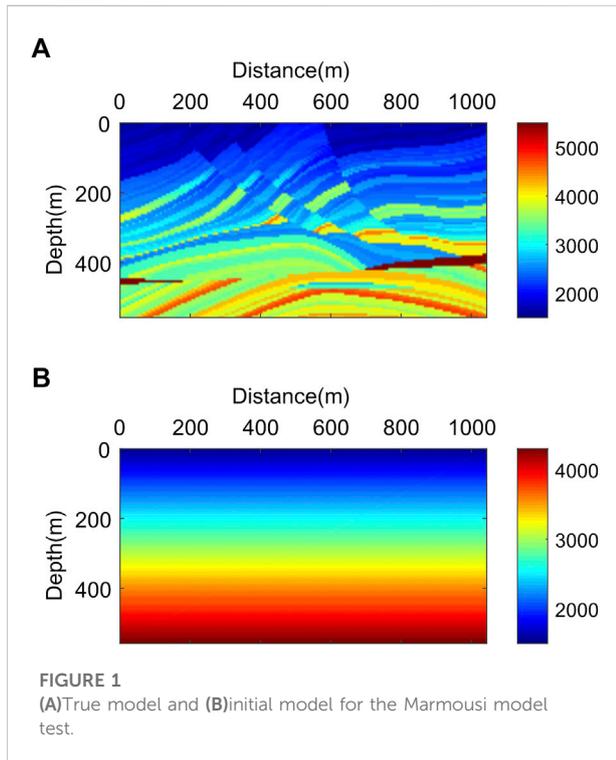
Full waveform inversion (FWI) is a non-linear optimization problem based on full-wavefield modeling to obtain quantitative information of subsurface structure by minimizing the difference between the observed seismic data and the predicted wavefield. The limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method is an effective quasi-Newton method in FWI due to its high inversion efficiency with low calculation and storage requirements. Like other conventional quasi-Newton methods, the approximation of the Hessian matrix in the L-BFGS method satisfies the quasi-Newton equation, which only exploits the gradient and model information while the available objective function value is neglected. The modified quasi-Newton equation considers the gradient, model, and objective function information together. Theoretical analysis reveals that the modified quasi-Newton equation is superior to the conventional quasi-Newton equation as it achieves higher-order accuracy in approximating the Hessian matrix. The modified L-BFGS method can be obtained by using the modified quasi-Newton equation to modify the L-BFGS method. This modification improves the accuracy of the Hessian matrix approximation with little increase of calculation for each iteration. We incorporate the modified L-BFGS method into FWI, numerical results show that the modified L-BFGS method has a higher convergence rate, achieves better inversion results, and has stronger anti-noise ability than the conventional L-BFGS method.

KEYWORDS

full waveform inversion, hessian matrix, quasi-Newton equation, modified quasi-Newton equation, L-BFGS, modified L-BFGS

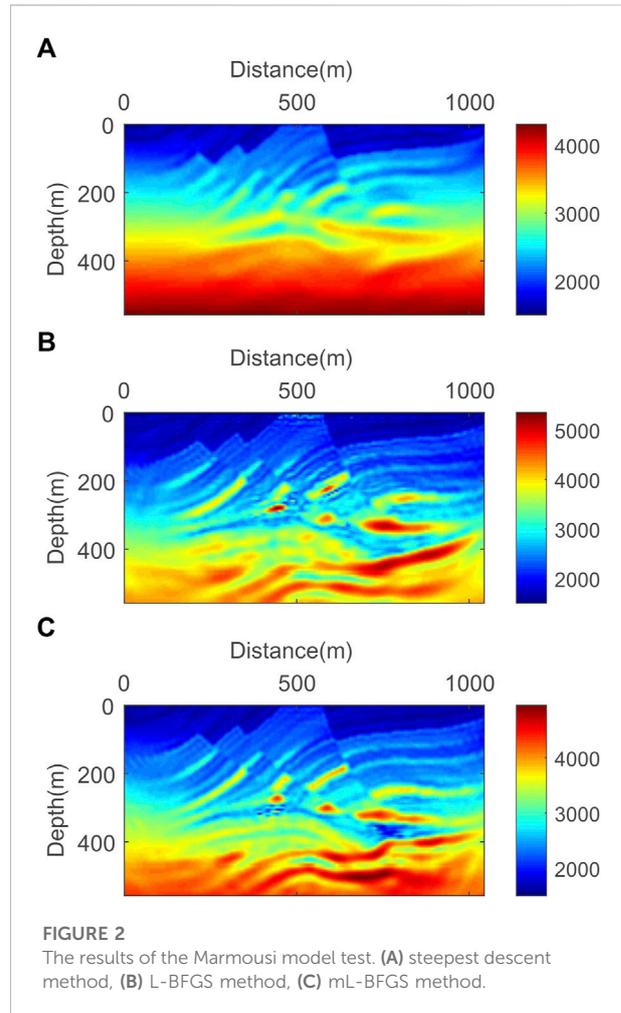
1 Introduction

Full waveform inversion (FWI) is a data-fitting method to extract quantitative information of subsurface structures by minimizing the residual between the observed seismic data and the predicted wavefield (Virieux and Operto, 2009). It is an important method to prospecting for oil- and gas-reservoirs. Frequency-domain FWI proposed by Pratt and Worthington (1990) uses an implicit finite difference scheme which is



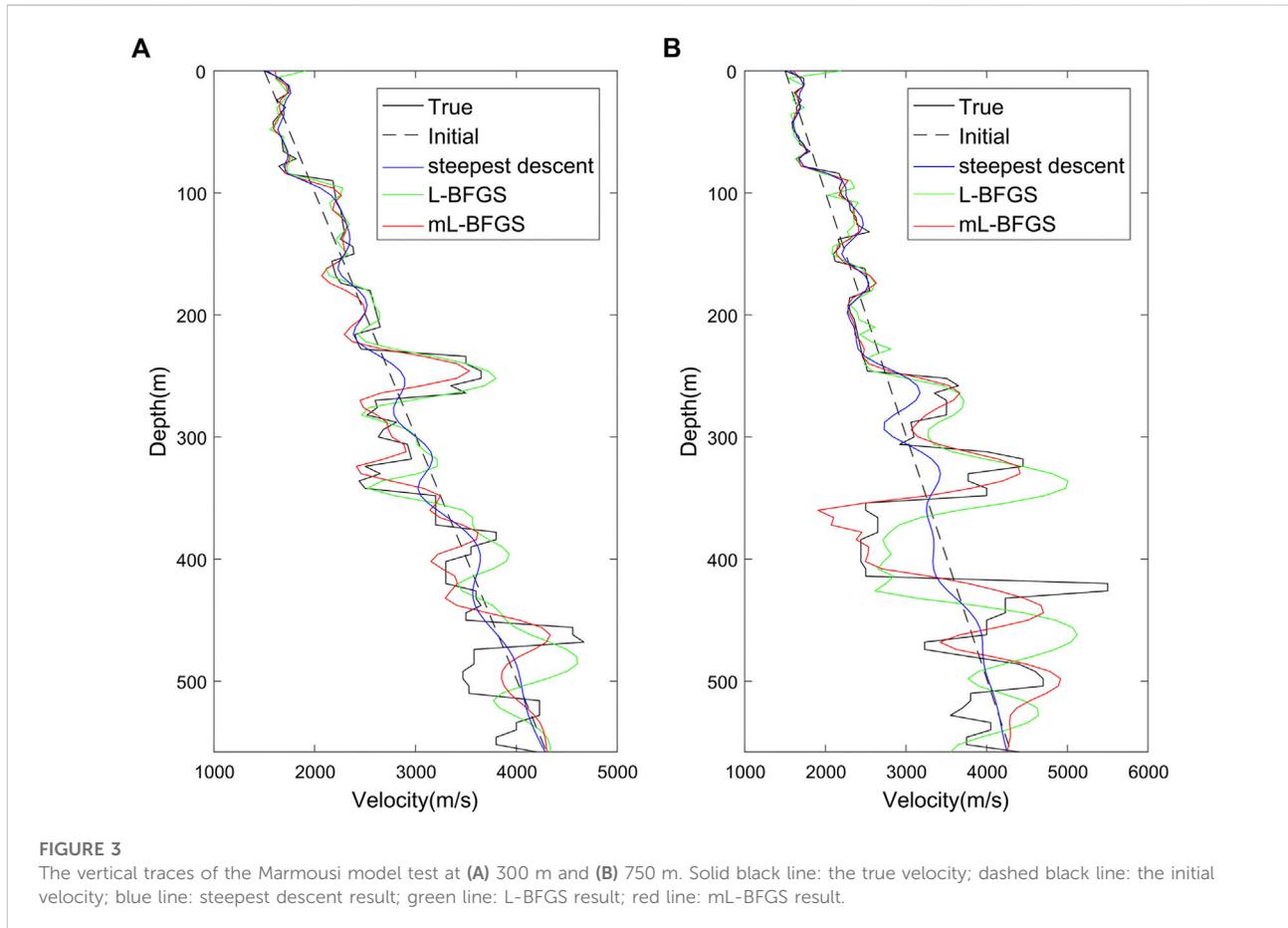
convenient for multi-shots computation. By successively iterating from low frequencies to high frequencies, FWI can more easily converge to the global minimum (Sirgue and Pratt, 2004). Time-domain FWI uses an explicit finite difference scheme and requires less memory in wavefield modeling (Tarantola, 1984). Bunks et al. (1995) proposed multiscale time-domain FWI to make the inversion stably converge to the global minimum through sequential iterations from low to high frequency scales.

Newton methods are effective methods for optimization, but they are not suitable for large-scale inverse problems due to the high computational cost of the Hessian matrix (Pratt et al., 1998). Instead of computing the Hessian matrix directly, the truncated Newton method computes the Newton descent direction by solving the corresponding linear system through an iterative procedure such as the conjugate gradient method (Metivier et al., 2013). Quasi-Newton methods construct an approximate inverse Hessian matrix by updating it over successive iterations (Nocedal and Wright, 1999). Many efforts have been made to improve the performance of quasi-Newton methods. Ma et al., 2010, Ma and Hale, 2011) used a projected Hessian matrix to reduce storage and computational costs. Wang et al. (2013) proposed a hybrid iterative optimization scheme, which can dynamically switch between inversion methods. Liu et al., 2015) applied a memoryless quasi-Newton method which can achieve acceptable result with low storage to FWI. Conventional quasi-Newton methods only exploit the gradient and model information while the available information of the objective



function value is ignored. Attempts have been made to consider the objective function information in optimization (Yuan, 1991; Yuan and Byrd, 1995) and modified quasi-Newton methods have been proposed (Wei et al., 2004, 2006). The modified quasi-Newton equation exploits the gradient, model, and objective function value information together, and it achieves higher-order accuracy in approximating the Hessian matrix with little increase in computational costs for each iteration (Zhang et al., 1999; Zhang and Xu, 2001). Liu and Liu, (2013) introduced the modified quasi-Newton equation (Zhang and Xu, 2001) into FWI, tested the performance of the modified Broyden-Fletcher-Goldfarb-Shanno (mBFGS) method on a simple model, and showed the superiority of the modified BFGS method over the conventional BFGS method.

However, the storage of the full dense approximate inverse Hessian matrix is still a challenge for large-scale problems. The limited-memory BFGS (L-BFGS) method is an adaptation of the BFGS method and a limited-memory quasi-Newton method widely-used in FWI. It reduces storage by constructing the approximate inverse Hessian matrix with several vector pairs of recent iterations instead of saving the entire matrix (Liu and



Nocedal, 1989). There are many different types of L-BFGS methods developed for FWI. Fabien-Ouellet et al. (2017) proposed a stochastic L-BFGS method that supports the stochastic random subsampling of sources. Dai et al. (2017) implemented a L-BFGS-based orthant-wise limited memory quasi-Newton method in l1-regularized FWI with prior model information. Liu et al. (2022) advanced L-BFGS and Hessian related pre-conditioners for uncertainty quantification in FWI and proposed a BFGS-RSVD workflow to achieve a faster Hessian retrieval. By applying the modification strategy to the L-BFGS method, one can obtain the modified L-BFGS (mL-BFGS) method, which considers the objective function information and more accurately approximates the inverse Hessian matrix than the conventional L-BFGS method (Yuan et al., 2010).

In this paper, we first review the derivation of the modified quasi-Newton equation, then introduce the modified L-BFGS method into FWI and give the pseudo code of the algorithm. Next, we compare the mL-BFGS method with the L-BFGS and steepest descent method through time-domain FWI on a regenerated Marmousi model. Then we compare the mL-BFGS method with the L-BFGS method through frequency-domain FWI on a regenerated Overthrust model with noise-

free and noise-added synthetic seismic data. Numerical results show that the mL-BFGS method only increases a small calculation amount in each iteration, converges faster, and achieves higher inversion accuracy with less computational resources. Moreover, the mL-BFGS method has stronger anti-noise ability than the L-BFGS method.

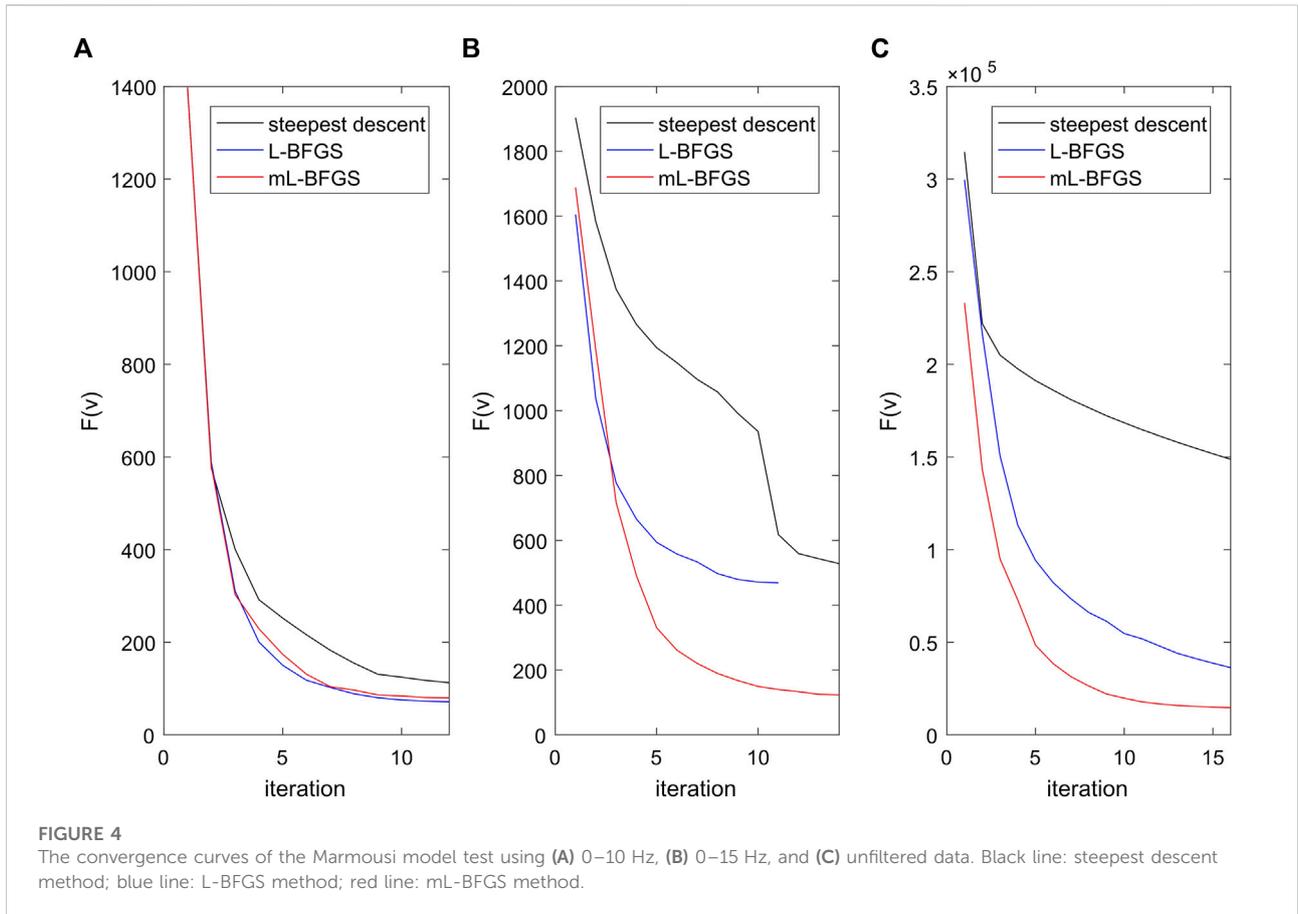
2 Theory

2.1 Quasi-Newton equation and modified quasi-Newton equation

Full waveform inversion is a non-linear optimization problem to minimize the objective function (Pratt et al., 1998)

$$F(\mathbf{v}) = \frac{1}{2} \delta \mathbf{d}^T \delta \mathbf{d}^*, \quad (1)$$

where \mathbf{v} represents the model parameters, superscripts T and * represent conjugate and transpose, respectively. $\delta \mathbf{d}$ is the misfit between the observed seismic data and the calculated seismic data. The objective function $F(\mathbf{v})$ is the least-squares norm of $\delta \mathbf{d}$.



One of the most important methods to solve this optimization problem is quasi-Newton method. It avoids the expensive calculation of the exact Hessian matrix by using the approximation of the Hessian matrix instead. The update formula of the quasi-Newton method is

$$\mathbf{v}_{k+1} = \mathbf{v}_k - \alpha_k \mathbf{H}_k \mathbf{g}_k, \tag{2}$$

where subscript k indicates the iteration number; α denotes the step length; \mathbf{g} represents the gradient of the objective function; and \mathbf{H} is an approximation of the inverse Hessian matrix that can be obtained from the quasi-Newton equation.

Then, we derive the quasi-Newton equation by second order Taylor expansion of the objective function $F(\mathbf{v})$ at \mathbf{v}_k

$$F(\mathbf{v}_k) = F(\mathbf{v}_{k+1}) - \mathbf{g}_{k+1}^T \mathbf{s}_k + \frac{1}{2} \mathbf{s}_k^T \mathbf{G}_{k+1} \mathbf{s}_k + O(\|\mathbf{s}_k\|^3), \tag{3}$$

where $\mathbf{s}_k = \mathbf{v}_{k+1} - \mathbf{v}_k$, and \mathbf{G} denotes the second-order derivative matrix of the objective function, i.e., \mathbf{G} is the Hessian matrix. We differentiate Eq. 3 with respect to the model parameter, and pre-multiply by \mathbf{s}_k^T to obtain

$$\mathbf{s}_k^T \mathbf{G}_{k+1} \mathbf{s}_k = \mathbf{s}_k^T \mathbf{y}_k + O(\|\mathbf{s}_k\|^3), \tag{4}$$

where $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$. Using \mathbf{B} to represent the approximation of Hessian matrix \mathbf{G} , we can rewrite Eq. 4 as

$$\mathbf{B}_{k+1} \mathbf{s}_k = \mathbf{y}_k. \tag{5}$$

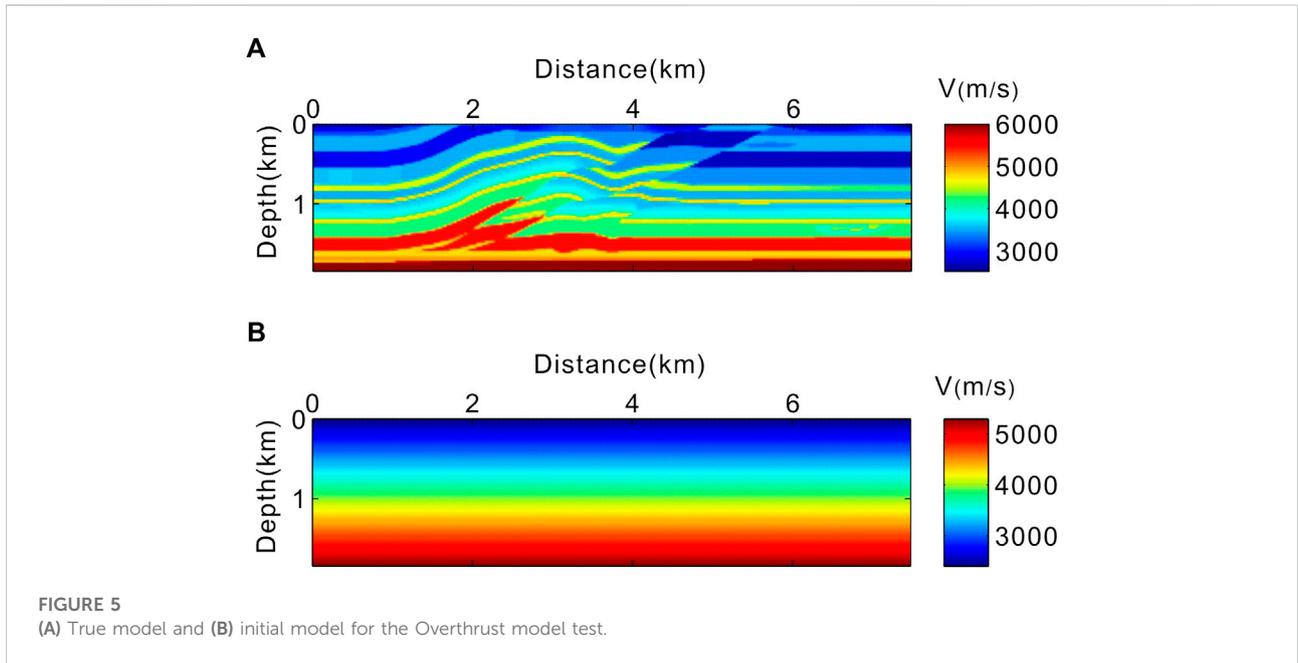
This is the quasi-Newton equation. The difference between the exact Hessian matrix and the approximate Hessian matrix \mathbf{B} can be expressed by the following formula:

$$\mathbf{s}_k^T \mathbf{B}_{k+1} \mathbf{s}_k = \mathbf{s}_k^T \mathbf{G}_{k+1} \mathbf{s}_k + O(\|\mathbf{s}_k\|^3), \tag{6}$$

As shown in Eq. 5, the quasi-Newton equation only exploits the gradient and model information, while the available objective function value information is neglected. In order to achieve higher-order accuracy in Hessian matrix approximation, Zhang et al. (1999) introduced the function value information to the quasi-Newton equation and proposed a modified quasi-Newton equation.

According to the example of Zhang et al. (1999), we derive the modified quasi-Newton equation through third-order Taylor expansion of the objective function $F(\mathbf{v})$ at \mathbf{v}_k

$$F(\mathbf{v}_k) = F(\mathbf{v}_{k+1}) - \mathbf{g}_{k+1}^T \mathbf{s}_k + \frac{1}{2} \mathbf{s}_k^T \mathbf{G}_{k+1} \mathbf{s}_k - \frac{1}{6} \mathbf{s}_k^T (\mathbf{T}_{k+1} \mathbf{s}_k) \mathbf{s}_k + O(\|\mathbf{s}_k\|^4), \tag{7}$$



where \mathbf{T} denotes the third order derivative matrix of the objective function. We differentiate Eq. 7 with respect to the model parameter, and pre-multiply \mathbf{s}_k^T to obtain

$$\mathbf{s}_k^T \mathbf{g}_k = \mathbf{s}_k^T \mathbf{g}_{k+1} - \mathbf{s}_k^T \mathbf{G}_{k+1} \mathbf{s}_k + \frac{1}{2} \mathbf{s}_k^T (\mathbf{T}_{k+1} \mathbf{s}_k) \mathbf{s}_k + O(\|\mathbf{s}_k\|^4), \quad (8)$$

Substitute Eq. 8 into Eq. 7 and eliminate \mathbf{T}_{k+1} to obtain

$$\mathbf{s}_k^T \mathbf{G}_{k+1} \mathbf{s}_k = \mathbf{s}_k^T \mathbf{y}_k + 6[F(\mathbf{v}_k) - F(\mathbf{v}_{k+1})] + 3(\mathbf{g}_k + \mathbf{g}_{k+1})^T \mathbf{s}_k + O(\|\mathbf{s}_k\|^4), \quad (9)$$

We use $\hat{\mathbf{B}}$ to represent the approximation of Hessian matrix \mathbf{G} , and rewrite Eq. 9 to obtain the modified quasi-Newton equation:

$$\hat{\mathbf{B}}_{k+1} \mathbf{s}_k = \hat{\mathbf{y}}_k, \quad (10)$$

where

$$\hat{\mathbf{y}}_k = \mathbf{y}_k + \frac{\theta_k}{\mathbf{s}_k^T \mathbf{z}} \mathbf{z}, \quad (\mathbf{s}_k^T \mathbf{z} \neq 0), \quad (11)$$

$$\theta_k = 6[F(\mathbf{v}_k) - F(\mathbf{v}_{k+1})] + 3(\mathbf{g}_k + \mathbf{g}_{k+1})^T \mathbf{s}_k. \quad (12)$$

In this paper, the vector \mathbf{z} in Eq. 11 is chosen to be \mathbf{s}_k to ensure that the denominator is not zero. As shown in Eqs. 10–12, the modified quasi-Newton equation considers the information of gradient, model, and objective function at the same time, while the calculation of vector addition and multiplication only increases a little. The difference between the exact Hessian matrix and the approximate Hessian matrix $\hat{\mathbf{B}}$ can be expressed as:

$$\mathbf{s}_k^T \hat{\mathbf{B}}_{k+1} \mathbf{s}_k = \mathbf{s}_k^T \mathbf{G}_{k+1} \mathbf{s}_k + O(\|\mathbf{s}_k\|^4), \quad (13)$$

Comparing Eq. 6, 13, we find that the modified quasi-Newton equation is superior to the conventional quasi-

Newton equation as it achieves higher-order accuracy in approximating the Hessian matrix.

2.2 Modified L-BFGS method

For large optimization problems with many variables, the approximate inverse Hessian matrix is usually dense, so the calculation and storage requirements of quasi-Newton methods like the BFGS method are excessive. The L-BFGS method is widely-used to solve large-scale problems while requiring less storage. Instead of storing the entire matrix, the L-BFGS method only stores a certain number of vector pairs $\{\mathbf{s}_k, \mathbf{y}_k\}$, and implicitly forms the inverse Hessian matrix (Nocedal and Wright, 1999).

The update formula of the modified BFGS (mBFGS) method is

$$\mathbf{H}_{k+1}^{MBFGS} = \mathbf{V}_k^T \mathbf{H}_k^{MBFGS} \mathbf{V}_k + \rho_k \mathbf{s}_k \mathbf{s}_k^T, \quad (14)$$

where

$$\rho_k = \frac{1}{\hat{\mathbf{y}}_k^T \mathbf{s}_k}, \mathbf{V}_k = \mathbf{I} - \rho_k \hat{\mathbf{y}}_k \mathbf{s}_k^T, \quad (15)$$

\mathbf{I} denotes the identity matrix, and ρ_k equals $\frac{1}{\hat{\mathbf{y}}_k^T \mathbf{s}_k}$. Yuan et al. (2010) gave the modified L-BFGS method, which only stores the information of several recent iterations:

$$\begin{aligned} \mathbf{H}_k^{ML-BFGS} = & (\mathbf{V}_{k-1}^T \dots \mathbf{V}_{k-m}^T) \mathbf{H}_k^{ML-BFGS(0)} (\mathbf{V}_{k-m} \dots \mathbf{V}_{k-1}) \\ & + \rho_{k-m} (\mathbf{V}_{k-1}^T \dots \mathbf{V}_{k-m+1}^T) \mathbf{s}_{k-m} \mathbf{s}_{k-m}^T (\mathbf{V}_{k-m+1} \dots \mathbf{V}_{k-1}) \\ & + \rho_{k-m+1} (\mathbf{V}_{k-1}^T \dots \mathbf{V}_{k-m+2}^T) \mathbf{s}_{k-m+1} \mathbf{s}_{k-m+1}^T (\mathbf{V}_{k-m+2} \dots \mathbf{V}_{k-1}) \\ & + \dots + \rho_{k-1} \mathbf{s}_{k-1} \mathbf{s}_{k-1}^T, \end{aligned} \quad (16)$$

In Eq. 16, the recent m vector pairs $\{\mathbf{s}_k, \hat{\mathbf{y}}_k\}$ are used, and the information of model, gradient, and objective function are considered together. The initial inverse Hessian matrix $\mathbf{H}^{ML-BFGS(0)}$ is constructed by using the pseudo-Hessian matrix (Shin et al., 2001).

Now we present the pseudo-code for the mL-BFGS algorithm in FWI as follows:

```

q = gk;
for  $i = k-1, k-2, \dots, k-m$ 
     $a_i = \rho_i \mathbf{s}_i^T \mathbf{q}; \quad (\rho_i = \frac{1}{\mathbf{y}_i^T \mathbf{s}_i})$ 
     $\mathbf{q} = \mathbf{q} - a_i \hat{\mathbf{y}}_i;$ 
end
r =  $\mathbf{H}_k^{MBFGS(0)} \mathbf{q};$ 
for  $i = k-m, k-m+1, \dots, k-1$ 
     $\beta = \rho_i \hat{\mathbf{y}}_i^T \mathbf{r};$ 
     $\mathbf{r} = \mathbf{r} + \mathbf{s}_i (a_i - \beta);$ 
end
stop with result  $\mathbf{r} = \mathbf{H}_k^{MBFGS} \mathbf{g}_k$  .

```

Algorithm 1

Algorithm 2

1. Get initial model \mathbf{v}_0 ; integer $m > 0, k = 0$;
2. $k = k + 1$; compute $\mathbf{s}_k, \mathbf{y}_k, \theta_k$;
Compute $\hat{\mathbf{y}}_k = \mathbf{y}_k + \frac{\theta_k}{\mathbf{s}_k^T \mathbf{z}}$;
3. **if** $k > m$
Discard the vector pair $\{\mathbf{s}_{k-m}, \hat{\mathbf{y}}_{k-m}\}$ from storage;
Save $\{\mathbf{s}_k, \hat{\mathbf{y}}_k\}$;
4. Give the initial approximate inverse Hessian matrix $\mathbf{H}_k^{MBFGS(0)}$;
5. Compute the update direction with Algorithm 1;
Get the step length through line search and compute the updated model:
 $\mathbf{v}_{k+1} = \mathbf{v}_k - \alpha_k \mathbf{r}$;
6. **if** termination condition satisfied
stop with \mathbf{v}_{k+1} ;
else
go to step 2.

3 Numerical examples

3.1 Time-domain FWI on marmousi model

In this section, we incorporate the mL-BFGS method into time-domain FWI, then compare the results with the conventional L-BFGS method and steepest descent method. A modified Marmousi model is generated by resampling a

representative region of the Marmousi model and is used as the true velocity model, as shown in Figure 1A. The depth and length of the model are 558 and 1,044 m, respectively, and the grid interval is 6 m. There are 175 receivers laid on the surface with a spatial interval of 6 m, and 25 shots located at a depth of 6 m with a spatial interval of 30 m. The initial model is a laterally homogeneous model with velocity linearly increasing from 1,500 m/s at the surface to 4,300 m/s at the bottom (Figure 1B).

A Ricker wavelet with a main frequency of 30 Hz is used. We consider cutoff frequencies of 10 and 15 Hz and perform low-pass filtering on the Ricker wavelet. By sequentially using the 0–10 and 0–15 Hz low-pass filtered Ricker wavelet and unfiltered Ricker wavelet as the source, we perform multiscale FWI from low to high frequency (Bunks et al., 1995) and set the maximum iteration numbers of the three different scales to 12, 14, and 16, respectively. The termination condition is $|\frac{F(\mathbf{v}_{k+1}) - F(\mathbf{v}_k)}{F(\mathbf{v}_k)}| \leq 0.005$.

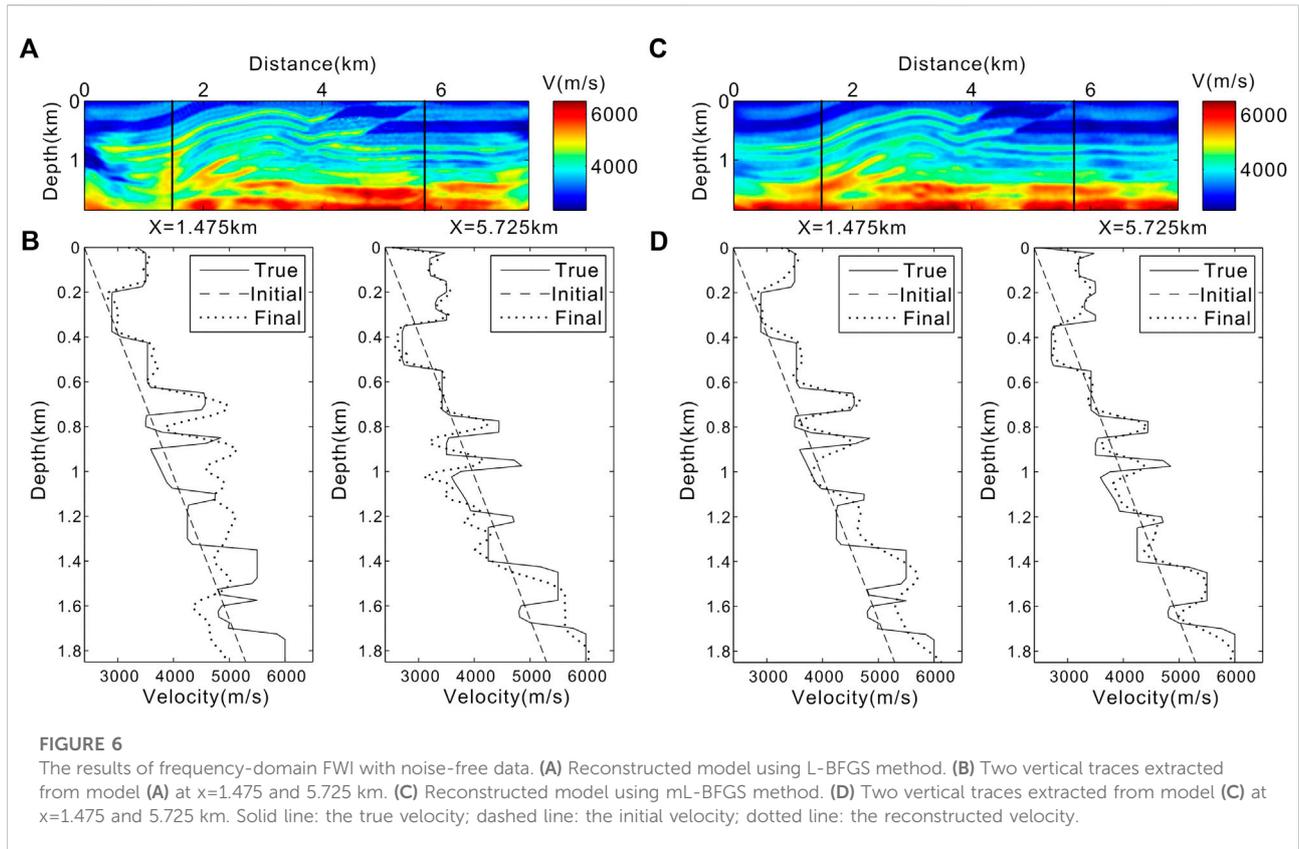
The multiscale FWI results of the steepest descent, L-BFGS and mL-BFGS methods are shown in Figure 2. The velocity structures at the deep part of the model are not well recovered in the result of steepest descent method (Figure 2A), while the inversion using the L-BFGS (Figure 2B) and mL-BFGS (Figure 2C) method rebuilds this part better. To make further comparison, two velocity traces at $x = 300$ m (Figure 3A) and $x = 750$ m (Figure 3B) are extracted from the reconstructed models. We can find that the velocity traces obtained by the mL-BFGS method are more consistent with the true velocity traces than the velocity traces obtained using L-BFGS method.

Figure 4 shows the comparison of the convergence curves of the steepest descent, L-BFGS and mL-BFGS methods in the three scales. It can be seen that the convergence rate of steepest descent method is slower than the L-BFGS and mL-BFGS methods.

3.2 Frequency-domain FWI on overthrust model

3.2.1 Noise-free data

In this section, we incorporate the mL-BFGS method into the frequency-domain FWI, and compare the results with the conventional L-BFGS method. A modified Overthrust model is generated by resampling a representative region of the Overthrust model and is used as the true velocity model, as shown in Figure 5A. The depth and length of the model are 1.875 and 7.5 km, respectively, and the grid interval is 25 m. There are 75 receivers laid on the surface with a spatial interval of 100 m, and 74 shots are located at a depth of 25 m with a spatial interval of 100 m. The initial model is a laterally homogeneous model, where the velocity increases linearly from 2,400 m/s at the surface to 5,296 m/s at the bottom (Figure 5B).



We generate the synthetic data by an average-derivative optimal frequency-domain modeling algorithm (Chen, 2012). A Ricker wavelet with a central frequency of 10 Hz is used. We select seven frequencies in the inversion process: 2.7, 3.7, 4.9, 7.1, 10.0, 14.2, and 20.0 Hz. The inversion iterations are performed from low frequency to high frequency in sequence, and jump into the next frequency when 20 iterations are completed or the termination condition $|\frac{F(v_{k+1})-F(v_k)}{F(v_k)}| \leq 0.005$ is satisfied.

Figure 6A shows the frequency-domain FWI result using the L-BFGS method. The velocity structures of the left deep part of the model are not well recovered, while the inversion using the mL-BFGS method rebuilds this part better (Figure 6C). To make further comparison, two velocity traces at x=1.475 km and x=5.725 km are extracted from the reconstructed models using the L-BFGS (Figure 6B) and mL-BFGS (Figure 6D) methods,

respectively. It can be found that the velocity traces obtained by the mL-BFGS method are more consistent with the true velocity traces than the velocity traces obtained using L-BFGS at the depth of 0.7 km–1.875 km.

Ben-Hadj-Ali et al. (2011) proposed a factor to assess the error of inversion by quantitatively evaluate the differences between the true model and the reconstructed model using the following formula

$$err = \left(\frac{\|v_{inv} - v_{true}\|^2}{\|v_{true}\|^2} \right)^{1/2} \times 100\%, \quad (17)$$

where v_{inv} and v_{true} denote the recovered model and true model, respectively. According to Eq. 17, the error factor of rebuilt models using L-BFGS and mL-BFGS method is 10.63% and 7.56%, respectively (Table 1), which illustrates that the

TABLE 1 Comparison between L-BFGS and mL-BFGS in FWI with noise-free data.

| Methods | Consuming time (s) | Iteration number | Average consuming time each iteration (s) | Error (%) |
|---------|--------------------|------------------|-------------------------------------------|-----------|
| L-BFGS | 5,804.42 | 114 | 50.92 | 10.63 |
| mL-BFGS | 2,802.65 | 55 | 50.96 | 7.56 |

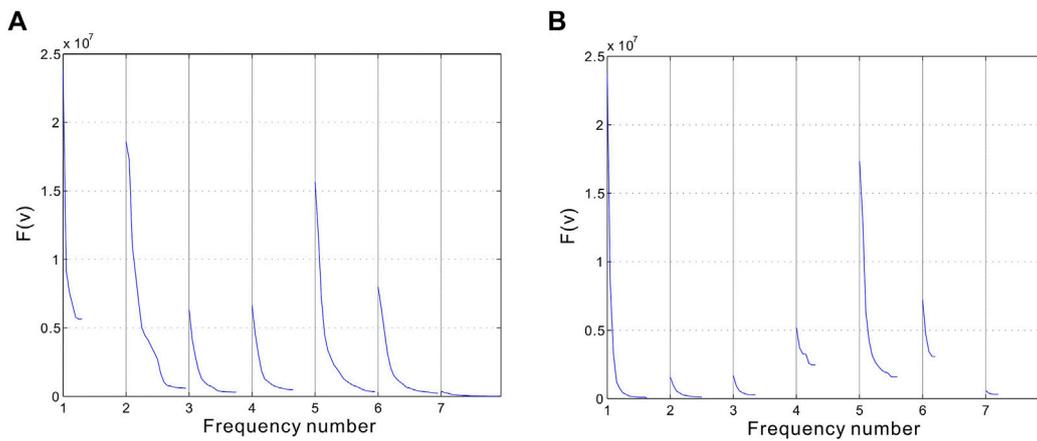


FIGURE 7
The convergence curves of (A) L-BFGS method and (B) mL-BFGS method of the Overthrust model test with noise-free data.

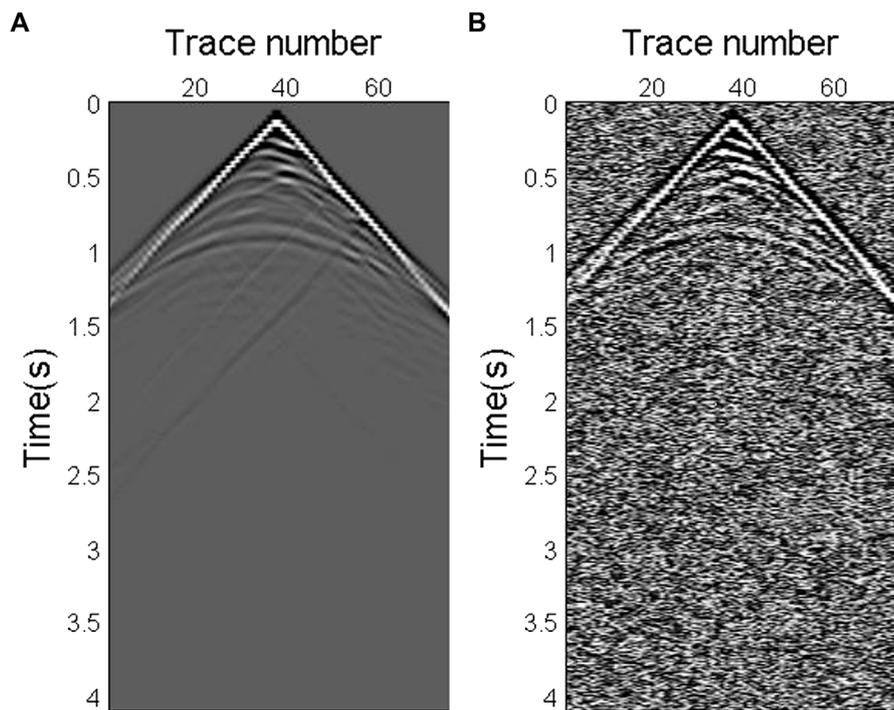


FIGURE 8
(A) The original seismogram. (B) The seismogram contaminated by Gaussian noise.

inversion result using mL-BFGS method is less different from the true model and more accurate.

The convergence curves of the L-BFGS and mL-BFGS methods are shown in Figures 7A,B, respectively. The vertical axis denotes the value of the misfit function, and

the horizontal axis denotes the 7 frequency numbers. The total iteration number of L-BFGS is 114, while the total iteration number of mL-BFGS is only 55. As shown in Table 1, the average calculation time for each iteration of the two methods is almost the same, but the mL-BFGS method converges faster

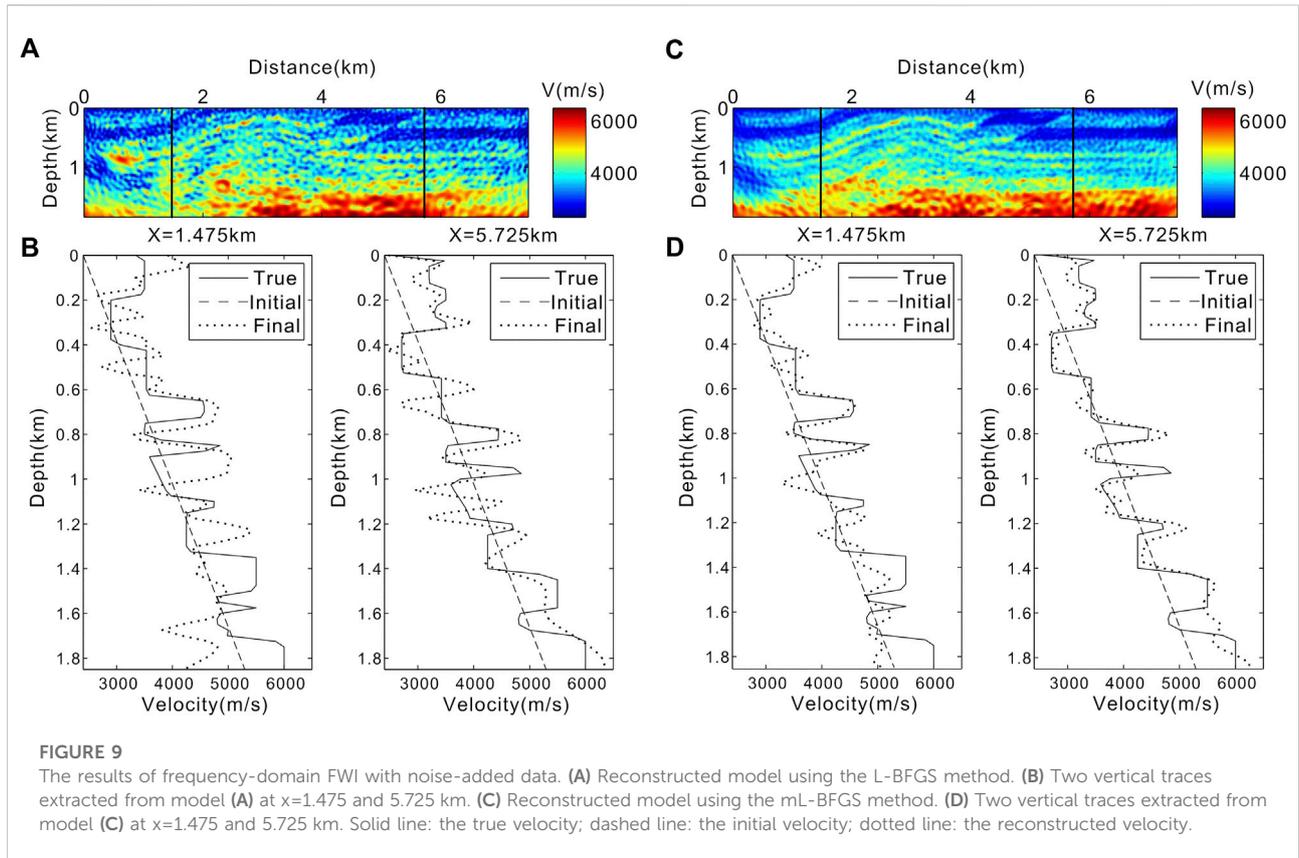


TABLE 2 Comparison between L-BFGS and mL-BFGS in FWI with noise-added data.

| Methods | Consuming time (s) | Iteration number | Average consuming time each iteration (s) | Error (%) |
|---------|--------------------|------------------|-------------------------------------------|-----------|
| L-BFGS | 4,224.22 | 83 | 50.89 | 14.38 |
| mL-BFGS | 1732.19 | 34 | 50.95 | 10.58 |

and requires fewer iterations. Therefore, the mL-BFGS method obtains better inversion result than the L-BFGS method with less calculation time.

3.2.2 Noise-added data

Actual seismic data are always contaminated by noise. Therefore, we introduce noise into the synthetic seismic data, and study the performance of the mL-BFGS method with noisy data in this section. We construct the noisy synthetic data (Figure 8B) by introducing Gaussian noise to the original synthetic data (Figure 8A) using the suaddnoise procedure of Seismic Unix (Cohen and Stockwell, 2008), with the S/N parameter equal to 40. Except for the noise-added seismic data, there are no differences between the experiment settings of the numerical example in this section and the previous numerical example with noise-free data.

As illustrated in Figure 9A, the model reconstructed using the L-BFGS method is blurred and severely contaminated by artefacts. This phenomenon is reflected in the extracted velocity traces (Figure 9B), which show that the recovered velocity deviates from the true velocity significantly. Comparing with the model recovered using the L-BFGS method, the model reconstructed by the mL-BFGS method (Figure 9C) is less contaminated by artefacts, and the model layers are more continuous. As shown in Figure 9D, the extracted velocity traces are in better accordance with the true velocity than those in Figure 9B. According to Eq. 17, the error factor of rebuilt models using L-BFGS and mL-BFGS method with noise-added data is 14.38% and 10.58%, respectively (Table 2).

The convergence curves of the L-BFGS and mL-BFGS methods with noise-added data are shown in Figures 10A,B, respectively. The total number of iterations of the mL-BFGS

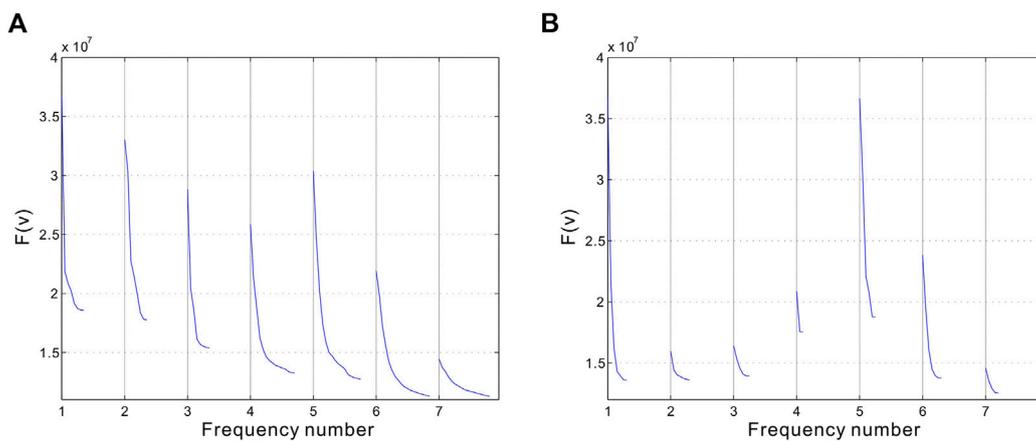


FIGURE 10

The convergence curves of (A) L-BFGS method and (B) mL-BFGS method of the Overthrust model test with noise-added data.

method is 34, which is markedly less than the total number of iterations, 83, of the L-BFGS method. It can be seen from Table 2 that with similar average calculation time for each iteration, the mL-BFGS method consumed less computing time in total because it converges faster and takes fewer iterations than the L-BFGS method, and the error factor for the reconstructed model with the mL-BFGS method is smaller than that with L-BFGS.

4 Conclusion and perspectives

We incorporated the modified L-BFGS method into full waveform inversion. The modified L-BFGS method considers the gradient, model, and function information together, and achieves higher-order accuracy for approximating the inverse Hessian matrix than the conventional L-BFGS method, while calculation time does not increase significantly for each iteration. Through numerical experiments incorporating modified L-BFGS into time-domain FWI on a regenerated Marmousi model and frequency-domain FWI on a regenerated Overthrust model with noise-free and noise-added synthetic seismic data, the modified L-BFGS method shows some advantages over the L-BFGS method including: higher convergence speed, less computation time, better inversion results, and stronger anti-noise ability. Therefore, the modified L-BFGS is an effective method in full waveform inversion.

In this paper, we only considered conventional FWI with simple L2-norm objective function, while there are other types of FWI based on different misfit functions, like the envelope inversion, travelt ime inversion, and FWI using the deconvolution-based objective function. It is theoretically feasible to incorporate the

modified L-BFGS method into these inversion methods with other types of misfit function.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

M-XD and B-SH contributed to the idea and methodology, M-XD did the numerical tests and wrote this manuscript, and W-SH checked and polished the manuscript. All authors have read the manuscript and agree to publish it.

Funding

This work was supported by the Open Foundation of Engineering Research Center of Nuclear Technology Application, Ministry of Education (HJSJYB2017-7), the Science and Technology Research project of the Jiangxi Provincial Education Department (GJJ170481), and the National Natural Science Foundation of China (42274149).

Acknowledgments

The authors appreciate the editors and reviewers for their valuable comments and suggestions which greatly helped us to

improve the present manuscript. We would like to thank professor David Nobes for his help with the English expression.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Ben-Hadj-Ali, H., Operto, S., and Virieux, J. (2011). An efficient frequency-domain full waveform inversion method using simultaneous encoded sources. *Geophysics* 76 (4), R109–R124. doi:10.1190/1.3581357
- Bunks, C., Saleck, F. M., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics* 60 (5), 1457–1473. doi:10.1190/1.1443880
- Chen, J. B. (2012). An average-derivative optimal scheme for frequency-domain scalar wave equation. *Geophysics* 77 (6), T201–T210. doi:10.1190/geo2011-0389.1
- Cohen, J. K., and Stockwell, J. J. W. (2008). *CWP/SU: Seismic Unix release no.41: An open source software package for seismic research and processing*. United States: Center for Wave Phenomena, Colorado School of Mines.
- Dai, M. X., Chen, J. B., and Cao, J. (2017). ℓ_1 -Regularized full waveform inversion with prior model information based on orthant-wise limited memory quasi-Newton method. *J. Appl. Geophys.* 142, 49–57. doi:10.1016/j.jappgeo.2017.03.020
- Fabien-Ouellet, G., Gloaguen, E., and Giroux, B. (2017). “Expanded abstracts: Seg,” in *A stochastic L-BFGS approach for full waveform inversion: 87th annual international meeting* (United States: SEG Library), 1622–1627.
- Liu, C., Gao, F., Feng, X., Liu, Y., and Ren, Q. (2015). Memoryless quasi-Newton (MLQN) method for 2D acoustic full waveform inversion. *Explor. Geophys.* 46 (2), 168–177. doi:10.1071/eg13090
- Liu, D. C., and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Math. Program.* 45 (1), 503–528. doi:10.1007/bf01589116
- Liu, L., and Liu, H. (2013). “Expanded abstracts: Seg,” in *Waveform inversion with modified quasi-Newton algorithm: 83st annual international meeting* (United States: SEG Library), 3206–3210.
- Liu, Q., Beller, S., Lei, W., Peter, D., and Tromp, J. (2022). Pre-conditioned BFGS-based uncertainty quantification in elastic full waveform inversion. *Geophys. J. Int.* 228 (2), 796–815. doi:10.1093/gji/ggab375
- Ma, Y., and Hale, D. (2011). “Expanded abstracts: Seg,” in *A projected hessian matrix for full waveform inversion: 81st annual international meeting* (United States: SEG Library), 2401–2405.
- Ma, Y., Hale, D., Meng, Z., and Gong, B. (2010). Expanded abstracts: Seg, 1003–1007. *Full waveform inversion with image-guided gradient: 80th annual international meeting*, SEG Library, United States.
- Métivier, L., Brossier, R., Virieux, J., and Operto, S. (2013). Full waveform inversion and the truncated Newton method. *SIAM J. Sci. Comput.* 35 (2), B401–B437. doi:10.1137/120877854
- Nocedal, J., and Wright, S. J. (1999). *Numerical optimization*. New York, NY: Springer.
- Pratt, R. G., Shin, C., and Hicks, G. J. (1998). Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophys. J. Int.* 133 (2), 341–362. doi:10.1046/j.1365-246X.1998.00498.x
- Pratt, R. G., and Worthington, M. H. (1990). Inverse theory applied to multi-source cross-hole tomography. Part I: Acoustic wave-equation method. *Geophys. Prospect.* 38 (3), 311–329. doi:10.1111/j.1365-2478.1990.tb01847.x
- Shin, C., Jang, S., and Min, D. J. (2001). Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophys. Prospect.* 49 (5), 592–606. doi:10.1046/j.1365-2478.2001.00279.x
- Sirgue, L., and Pratt, R. G. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics* 69 (1), 231–248. doi:10.1190/1.1649391
- Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics* 49 (8), 1259–1266. doi:10.1190/1.1441754
- Virieux, J., and Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. *Geophysics* 74 (6), WCC1–WCC26. doi:10.1190/1.3238367
- Wang, Y., Dong, L. G., and Liu, Y. Z. (2013). Improved hybrid iterative optimization method for seismic full waveform inversion. *Appl. Geophys.* 10 (3), 265–277. doi:10.1007/s11770-013-0389-x
- Wei, Z., Li, G., and Qi, L. (2006). New quasi-Newton methods for unconstrained optimization problems. *Appl. Math. Comput.* 175 (2), 1156–1188. doi:10.1016/j.amc.2005.08.027
- Wei, Z. X., Yu, G. H., Yuan, G. L., and Lian, Z. G. (2004). The superlinear convergence of a modified BFGS-type method for unconstrained optimization. *Comput. Optim. Appl.* 29 (3), 315–332. doi:10.1023/b:coap.0000044184.25410.39
- Yuan, G., Wei, Z., and Wu, Y. (2010). Modified limited memory BFGS method with nonmonotone line search for unconstrained optimization. *J. Korean Math. Soc.* 47 (4), 767–788. doi:10.4134/jkms.2010.47.4.767
- Yuan, Y. X. (1991). A modified BFGS algorithm for unconstrained optimization. *IMA J. Numer. Anal.* 11 (3), 325–332. doi:10.1093/imanum/11.3.325
- Yuan, Y. X., and Byrd, R. H. (1995). Non-quasi-Newton updates for unconstrained optimization. *J. Comput. Math.* 13 (2), 95–107.
- Zhang, J., and Xu, C. (2001). Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations. *J. Comput. Appl. Math.* 137 (2), 269–278. doi:10.1016/s0377-0427(00)00713-5
- Zhang, J. Z., Deng, N. Y., and Chen, L. H. (1999). New quasi-Newton equation and related methods for unconstrained optimization. *J. Optim. Theory Appl.* 102 (1), 147–167. doi:10.1023/a:1021898630001

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.