# COMPRESSED FWI INVERSION WITH NON-MONOTONE LINE SEARCH LBFGS

CHAORAN DUAN[1], FENGJIAO ZHANG[1,2], LIGUO HAN[1], AO CHANG[1], XIAOCHUN YANG[1] and FEI HUANG[2]

[1] *College of Geo-Exploration Science and Technology, Jilin University, Changchun 130026, P.R. China. fengjiao.zhang@geo.uu.se*
[2] *Department of Earth Sciences, Uppsala University, Uppsala 75236, Sweden.*

(Received May 15, 2017; revised version accepted September 12, 2017)

ABSTRACT

Duan, C., Zhang, F., Han, L., Chang, X. and Yang, F., 2017. *Journal of Seismic Exploration*, 26: 561-585.

Full waveform inversion (FWI) is a high quality seismic imaging method. It is a nonlinear inversion problem which usually needs the monotone line search method to be solved. However, the speed of convergence for such a simple search technique is relatively slow. In this paper, we combine the non-monotone line search technique with the LBFGS method and apply them to the frequency-domain FWI. We test this new method on a two-dimensional Marmousi model. The results show that the method is robust. Comparing with the monotone line search method, the new method could improve the convergence rate of FWI. We also test the new method with a two-dimensional conventional streamer data set and the results show some improvements compared with the conventional FWI method.

KEY WORDS**:** full waveform inversion (FWI), LBFGS method, non-monotone line search, principal component analysis.

## INTRODUCTION

Full waveform inversion (FWI) could provide a high-quality model and help to obtain a better image in the complicated geologic structures. It is a data-fitting procedure based on the amplitude and phase information of seismic data. We define the $l_2$ norm of the difference between the calculated wave-field and observed wave-field as the objective function, then update the velocity model by minimizing it. The method was first introduced by Tarantola (1984) and developed by many others later (Benhadjali, 2011; Brossier, 2011; Schiemenz, 2013).

0963-0651/17/$5.00 

Tarantola proposed the time domain FWI based on the generalized least-square method. In his approach, the gradient was calculated by cross-correlating the forward wave-field and back-propagation residual wave-field which avoid to directly solve the Fréchet derivative matrix. Pratt (1999) proposed the frequency domain FWI in the 1990s which reduced the computational work by using less frequencies during the inversion.

Realization of FWI depends on the optimization methods such as the steepest-descent method, the Newton method, the quasi-Newton method, the conjugate-gradient method and so on. The quasi Newton method is less computational intensive than the Newton method since it only updates the approximation of the Hessian matrix at each iteration of the inversion, which avoids the calculation of the Fréchet derivative matrix. On the other hand, the quasi Newton method has inherited the advantages from the Newton method, which has a faster convergence speed compared with the steepest- descent method. The famous quasi Newton calibration formulas include R1(Broyden, 1965), SR1(Broyden , 1967) DFP，BFGS, PSB and so on. Among all the formulas above, The BFGS (Broyden–Fletcher–Goldfarb–Shanno algorithm) method is generally regarded as the most effective quasi Newton method for the optimization problems and has been widely used in the recent years (Ma, 2012; Cheng, 2010). It requires lower precision for line search when compared with the conjugated-gradient method (Byrd, 1989). However, the BFGS method has to store an N order matrix, which will cost lots of the internal storage with large model size. Necoda (1980) proposed the limited memory BFGS method（LBFGS method) which avoids to store the iterative matrix. The LBFGS method produces searching direction by utilizing the current negative gradient and m vector pairs $(s^{(i)}, y^{(i)})$ generated by the BFGS method. Here i = k − m + 1, … k. Thus, the LBFGS method can save the internal storage and also accepts the unit step-length. The most commonly used search technique at present is monotone line search. The monotone line search technique requires monotonically decreased values of the objective function, which reduces the convergence rate of the iterative sequence (Sun, 2002). Grippo (1986) first proposed the non-monotone line search technique in 1986 and provided a broader means for the inexact line search. It requires no monotonically decreased values of the objective function, resulting a more flexible step-length and faster convergence rate. It could also avoid the local minimum point during the iteration and overcome the Maratos effect (Panier, 1991). Han et. al. (1997) modified the non-monotone line search technique proposed by Grippo and demonstrated the global convergence of the BFGS method in aspect to the convex objective function under the modified non-monotone Wolfe search. There are also some new non-monotone line search techniques have been proposed by many others in this century such as Zhang et al. (2004) and Shi et al. (2006).

The application of FWI in field data is dramatically limited by the prohibitively high computational costs. To solve this problem, this paper adopt the principal component analysis (PCA) to reduce the model dimension (Liu et al., 2012). The main steps of the PCA method are singular value decomposition (SVD) of the misfit matrix, cumulative energy analysis and projection of the original data sets onto a lower-dimensional data space.

The non-monotone line search technique does not require the objective function to decrease monotonically but within M steps. This paper combined the non-monotone line search technique and LBFGS method in FWI. The new method is tested by using numerical examples based on the two-dimensional Marmousi model. The results show that FWI based on the non-monotone line search technique converge faster than the monotone line search FWI. The new method is also tested on a two-dimensional conventional streamer data. It improves the velocity model in the shallow parts compared with the conventional FWI method.

## METHOD

### FWI with principal component analysis

*Theory of FWI*

FWI is a data-fitting procedure which minimizes the residuals between the observed wave-field and calculated wave-field to achieve the physical parameters of underground medium. FWI is widely used as it could provide a high precision model and help to obtain a better image in the complicated geologic structures. In this study, we use the frequency domain acoustic FWI method. Here we summarize the method for the sake of completeness. The two-dimensional frequency domain acoustic wave equation can be expressed as below:

$$\nabla^2 u(\mathrm{x},\omega) + \frac{\omega^2}{v_p^2(\mathrm{x})} u(\mathrm{x},\omega) = -s(\mathrm{x},\omega) \quad , \tag{1}$$

where $v_P$ is the P-wave velocity, ω is the angle frequency， $u(x,\omega)$ and $s(x,\omega)$ are the wave-field in the frequency domain and the source term, respectively. Eq. (1) could be rewritten as a linear equation in the following matrix form:

$$A(\mathrm{x},\omega)\,\mathrm{u}(\mathrm{x},\omega) = s(\mathrm{x},\omega) \quad , \tag{2}$$

where $A(x,\omega)$ denotes the sparse impedance matrix. The seismic wave-field $u(x,\omega)$ can be achieved by solving eq. (2) given the velocity model and the source term.

We define the $l_2$ norm of the residual as an objective function:

$$E = \frac{1}{2}\|\delta D\|_2^2 = \frac{1}{2}\|d_{cal} - d_{obs}\|_2^2 \quad , \tag{3}$$

where $d_{\mathrm{obs}}$ and $d_{\mathrm{cal}}$ indicate the observed wave-field and the calculated wave-field. The updated model can be expressed as the sum of the initial model $m_0$ plus a perturbation model $\delta m$ under Born approximation.

$$m = m_0 + \delta m \quad , \tag{4}$$

$$\delta m = -H_a^{-1}\nabla E(m) \quad . \tag{5}$$

In eq. (5)， $H_a$ indicates the approximate Hessian matrix and $\nabla E(m)$ denotes the gradient of the objective function, which can be achieved by cross-correlation between the forward wave-field and the back-propagation residual wave-field:

$$\nabla E(m) = -\Re[(\frac{\partial A}{\partial m})^T d_{cal}(m)^T B^H] \quad . \tag{6}$$

FWI is carried out by the iterative inversion based on eqs. (4), (5) and (6).

*Principal component analysis*

Principal component analysis (PCA) is a classical mathematic and statistics method. It usually used in the data sets' analysis and compression (Moore, 1981; Abdi et al., 2010). It can reduce the data dimension and increasing the calculation speed. In general, it could be divided into three steps when applying to FWI (Liu et al., 2012). (1) singular value decomposition (SVD) of the misfit matrix, (2) cumulative energy analysis and (3) projection of the original data sets onto a lower-dimensional data space.

The singular value decomposition (SVD) of the misfit matrix $\delta\mathbf{D}$ in eq. (3) can be expressed as following

$$\delta\mathbf{D} = \mathbf{LXR}^{\mathrm{H}} \quad , \tag{7}$$

where **L** and **R** are complex unitary matrixes **X** is an $N_r \times N_s$ rectangular diagonal matrix of singular values $\sigma_i = \mathbf{X}_{\mathrm{i,i}}$. Then we compute the cumulative energy and select a suitable threshold $\chi$ to control the size of the projection matrix used for data reduction according to the equation below

$$\frac{\sum_{i=1}^{k}\sigma_i^2}{\sum_{i}\sigma_i^2} \geq \chi \quad . \tag{8}$$

We can receive a $N_s \times k$ projection matrix $\mathbf{R}_k$ by saving the first $k$ columns of matrix $\mathbf{R}$. The $k$ here in the PCA method is cross-ponding with source number. Projecting the original source matrix $\mathbf{S}$ observed data matrix $\boldsymbol{d}_{obs}$ and calculated data matrix $\boldsymbol{d}_{cal}$ onto the low-dimensional space using the projection matrix $\mathbf{R}_k$, we can get the dimension-reduced source matrix $\mathbf{S}_k = \mathbf{S}\mathbf{R}_k$, the observed data matrix $\boldsymbol{d}_{obsk} = \boldsymbol{d}_{obs}\mathbf{R}_k$ and the calculated data matrix $\boldsymbol{d}_{calk}$. Thus the objective function in eq. (3) can be written as:

$$E = \frac{1}{2}\left\| d_{calk} - d_{obsk} \right\|_2^2 \quad . \tag{9}$$

Replace eq. (3) by eq. (9) and iteratively update model $m$, we can get the final results of the FWI. Where the gradient is expressed as the following form after projecting into the low-dimensional space.

$$\nabla E(m) = -\Re[(\frac{\partial A}{\partial m})^T d_{calk}(m)^T B_k^H] \quad . \tag{10}$$

By using the principal component analysis method we can greatly increase the calculation speed without a great loss of inversion accuracy.

**Non-monotone line search**

The monotone line search technique requires the object function monotonically decreases which tends to reduce the convergence rate of iterative sequence. The non-monotone line search technique only requires the objective function to decrease within limited steps, which not only increases the convergence rate but also avoids the local minimum, thus overcome the Maratos effect. It has been widely used since 1980s (Grippo, 1989; Raydan, 1997; Zhu, 2003; Zhao, 2005; Huang, 2009).

Given the initial iteration point $x_0$ and the positive integer $m$, $M$, $m_0 = 0$, $\alpha > 0$, $0 < \rho < 1$, $\delta > 0$.

The non-monotone line search technique should satisfy the following condition:

$$\alpha_k = \alpha\rho^{h_k} \quad , \tag{11}$$

$$f(\mathrm{x}_k + \alpha_k d_k) \leq \max_{0 \leq j \leq m(k)} \{f(\mathrm{x}_{k-j})\} + \delta\alpha_k d_k^T g_k \quad , \tag{12}$$

where $h_k$ is the first non-negative integer which ensures step-length $a_k$ satisfies eq. （12, and $m(k) = \min\{ m(k-1) + 1, M\}$.

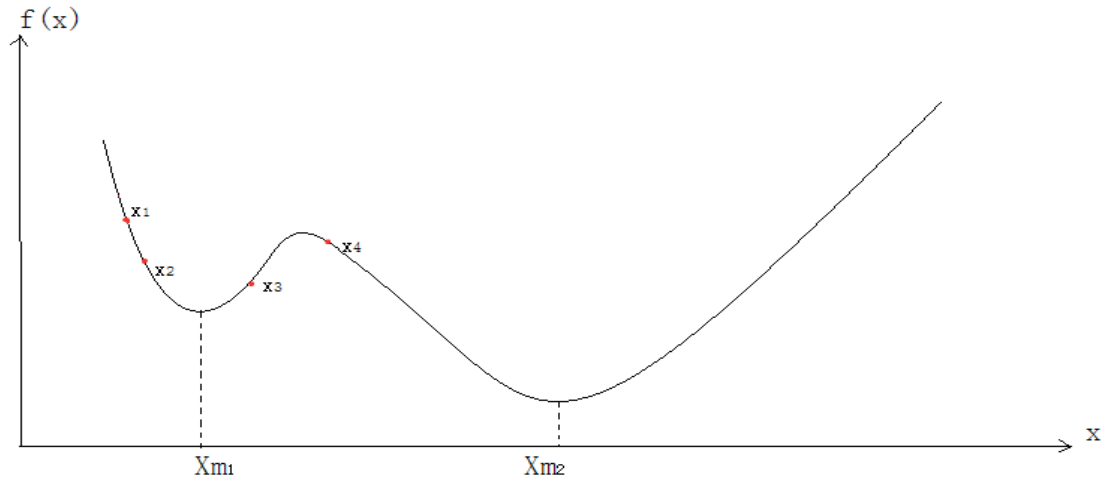



Fig. 1. Sketch for the non-monotone line search.

Fig. 1 shows an illustration of how the non-monotone line search technique may avoid the local minimum. $x$ denotes the model， $x_1$ is the initial iteration point, $f(x)$ indicates the objective function， $Xm_1$ is a local minimum point and $Xm_2$ is a global minimum point .The inversion may easily plunged into the local minimum point when using the monotone line search technique. This is because the monotone line search technique requires the values of the objective function to decrease monotonically, that is $f(x_1) > f(x_2) > f(x_3) > f(x_4)$. For the non-monotone line search technique only requires the values of the objective function to decrease in the limited steps, that is $\max\{f(x_1), f(x_2), f(x_3)\} > f(x_4)$. This allows the inversion to have more chance to avoid the local minimum point and reach the global minimum.

## FWI with NLBFGS

In this paper, we integrate the LBFGS method and the non-monotone line search technique into the FWI method. This will help to improve the ability of the FWI method in solving large-scale optimization problems and speeding up the convergence rate of the inversion. The inversion work-flow diagram is shown in Fig. 2. The whole work-flow could be divided into 5 steps:

(1) Choose initial parameters α， ρ， δ, $M$;
(2) Calculate model gradient $g_k$ based on the objective function;
(3) Determine step length α;
(4) Update the velocity model
(5) Repeat step 2 until the result satisfy the iteration stop criteria.

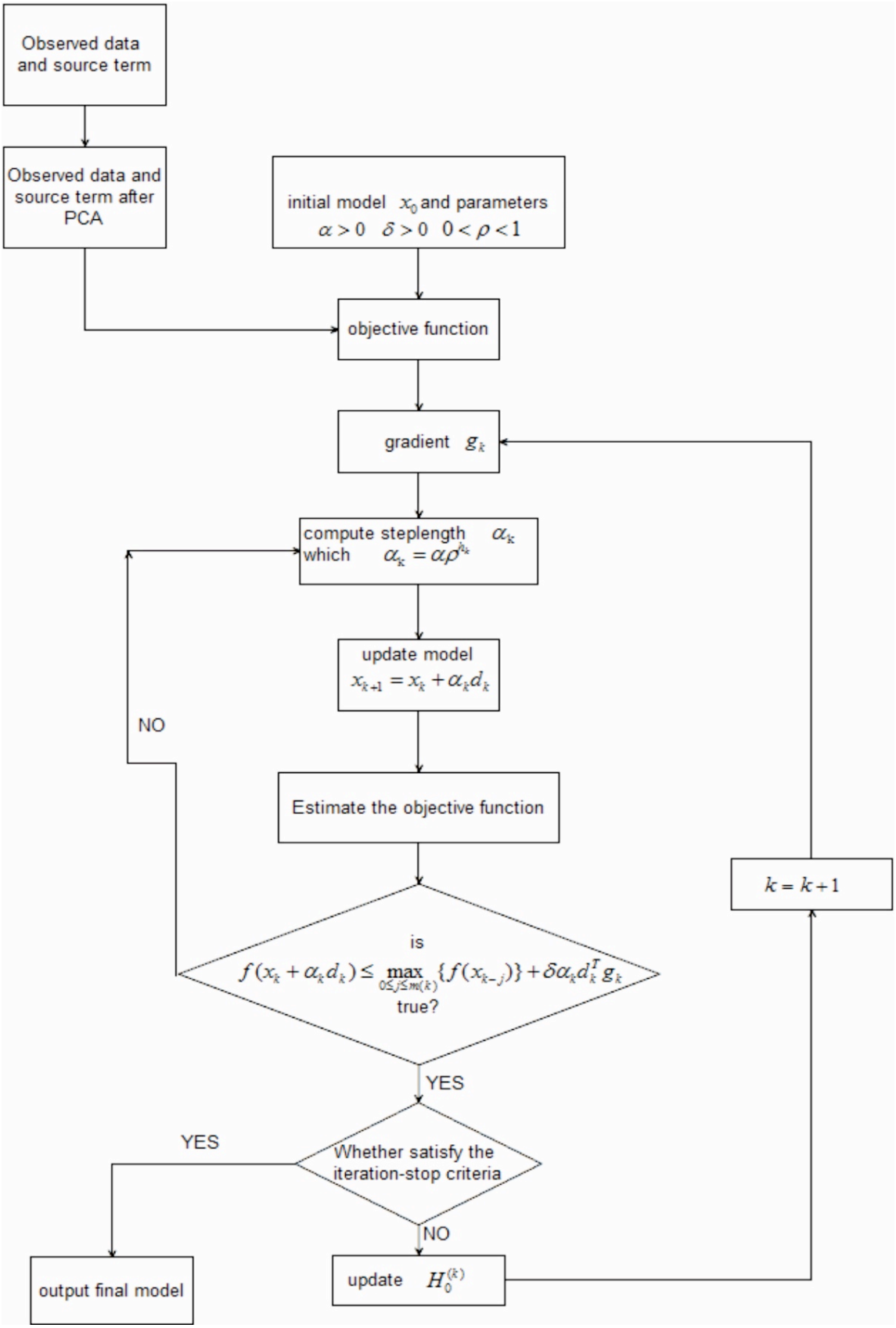


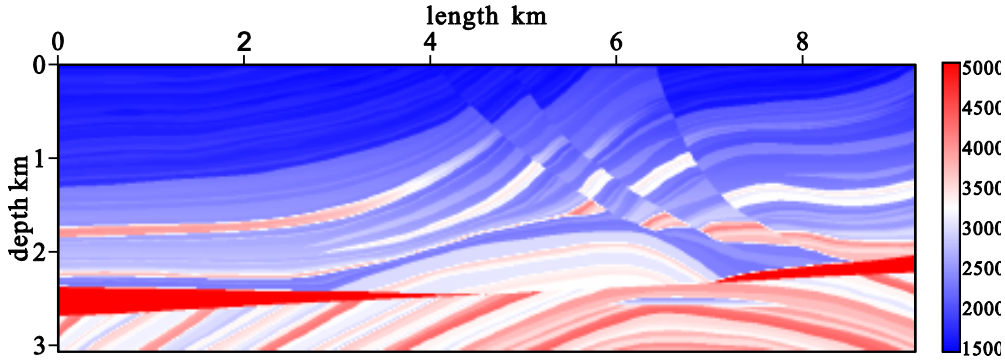

Fig. 2. The NLBFGS FWI work-flow.

## NUMERICAL EXAMPLE

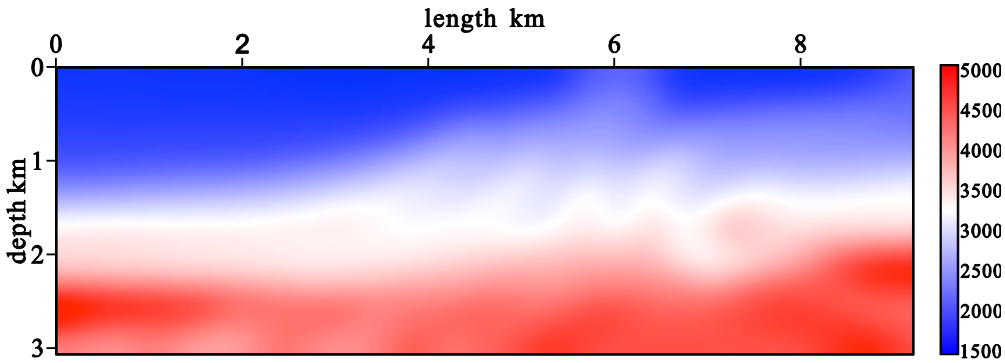
In this section, we use a synthetic data set to test this method. We also discuss the effects of inversion results by using different initial parameters.

### Synthetic data set and general inversion work flow

The synthetic data set is generated from a modified Marmousi model shown in Fig. 1(a). The speed range of this model is from 1500 m/s to 5000 m/s. The model size is 128×384 with 24 m gird size. The recording geometry consists 384 shots and 384 receivers with both 24 m shot and receiver interval. The source wavelet is 15 Hz Ricker wavelet. The range of used frequency in the inversion is from 2.9 to 27.9 Hz. We divide the inversion into two parts, lower and higher than 15 Hz. For the part lower than 15 Hz, we use the principal component analysis method to select 10 to 20 shots from the whole data set for the inversion. For the part higher than 15 Hz, we use 19 super shots with 480 m interval. The frequency has been divided into 10 groups with a 0.6 Hz interval and each group has 10 frequencies. Each group has been iterated for at least 25 times. The true model is shown in Fig. 3(a). At the beginning, the smooth version of the




(a) True model




(b) Initial starting model

Fig. 3. The initial starting velocity model.

true model is used as the starting model as shown in Fig. 3(b). The result from lower frequency group is used as the starting model for the higher frequency group.

## Initial parameters for the NLBFGS based FWI

The initial parameters play an important role in the NLBFGS base FWI. The choice of $\alpha, \rho, M$ will greatly affect the results of the inversion. We will discuss the effects of these parameters to the inversion results.

1. α is the initial step length and in theory the inversion only needs $\alpha > 0$. Here we take $\alpha = 0.1$, $\alpha = 2$ and $\alpha = 10$ as an example to discuss the influence of $\alpha$ on the inversion results.

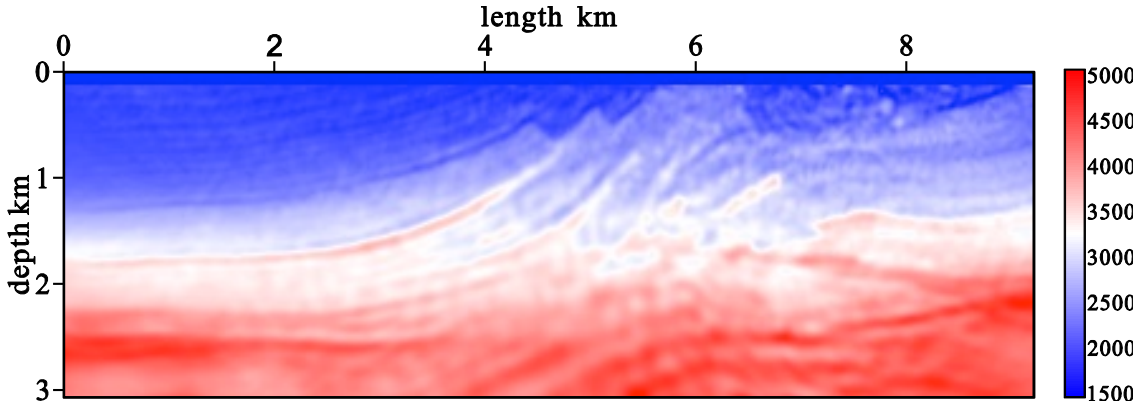
Table 1 illustrates some information of the NLBFGS FWI method with different $\alpha$. It shows the whole inversion time, total iteration number and the error between the results and true models. The precision of inversion is quantitatively expressed by error function

$$E_{res} = \frac{\left\| m_{inv} - m_{true} \right\|_2}{\left\| m_{true} \right\|_2} \times 100\% .$$

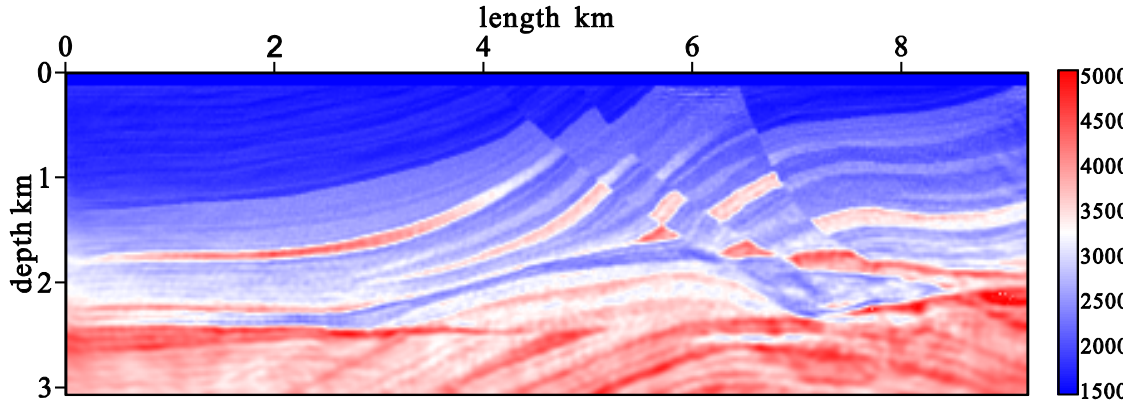
The smaller $E_{res}$, the higher the precision. In order to determine the impact of the value of $\alpha$ on the inversion results, we use the fixed value for M and $\rho$ ($M = 2, \rho = 0.2$). Fig. 4 shows the inversion results when different $\alpha$ is used. From Table 1 and Fig. 4 we could find that when $\alpha$ is small ($\alpha = 0.1$), $\alpha_k$ is also small. This makes the model update to be very slow and needs a lot of iterations to achieve relative good results. When α is large（α = 10), it needs to calculate many times to find a suitable $\alpha_k$. This also needs a lot time and reduces the convergence rate of the inversion. According to our tests，we find that we could get a relative reasonable results when α is between 1 and 5.

Table 1. The influence of $\alpha$ on the inversion results.

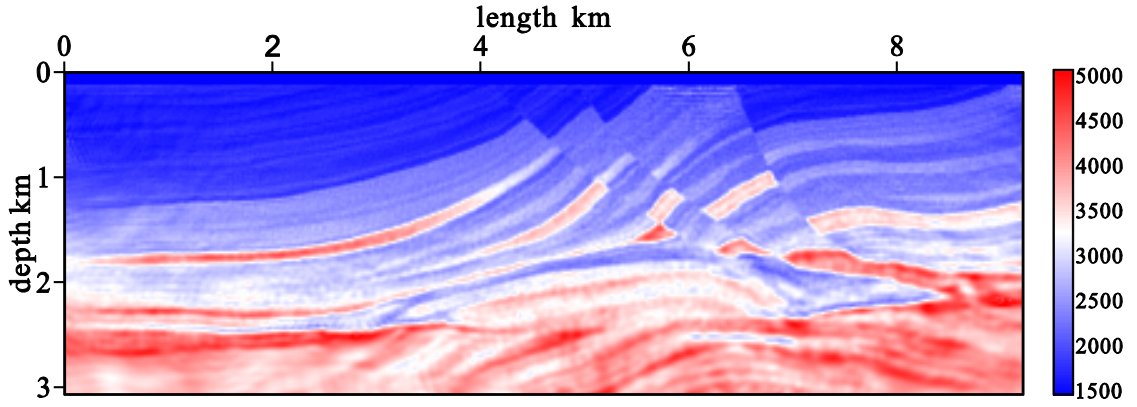
| $\alpha$ | Time(s) | Total iteration number | $E_{res}$(%) |
|---|---|---|---|
| $\alpha = 0.1$ | 161166 | 618 | 13.35 |
| $\alpha = 2$ | 119578 | 285 | 10.52 |
| $\alpha = 10$ | 157826 | 280 | 11.33 |




(a) Inversion result with $\alpha = 0.1$




(b) Inversion result with $\alpha = 2$




(c) Inversion result with $\alpha = 10$

Fig. 4. Inversion results with (a) $\alpha = 0.1$, (b) $\alpha = 2$, and (c) $\alpha = 10$.

2. $\rho$ is the attenuation coefficient of $\alpha$, in general we need 0 < ρ < 1. In this section, we discuss the influence of $\rho$ on the inversion results. Here we use $\rho = 0.01$, $\rho = 0.2$ and $\rho = 0.99$ as an example.

Table 2. The influence of $\rho$ on the inversion results.

| ρ | Time(s) | Total iteration number | $E_{res}(\%)$ |
|---|---|---|---|
| ρ = 0.01 | 157642 | 583 | 13.36 |
| ρ = 0.2 | 119578 | 285 | 10.52 |
| ρ = 0.99 | 239719 | 2500 | ------ |

Table 2 illustrates some information of the FWI method with different $\rho$. In the test we used fixed value for $M$ and $\alpha$ ($M = 2, \alpha = 2$) only $\rho$ is changed. Fig. 5 shows the inversion results with different $\rho$. From Table 2 and Fig. 5 we could find that when $\rho$ is very small ($\rho = 0.01$). As $\alpha_k = \alpha\rho^{h_k}$, $\alpha_k$ is also small. This will make the model update is very small for each iteration and reduce the calculation efficiency. When $\rho$ is too large ($\rho = 0.99$), $\alpha_k$ is almost unchanged during the iteration. If $\alpha_0$ did not satisfy the convergence condition, $\alpha_k$ will not update during the whole inversion. This will lead to a failure to the inversion. According to our test, we could get a relative good result when $\rho$ is between 0.1 and 0.7.

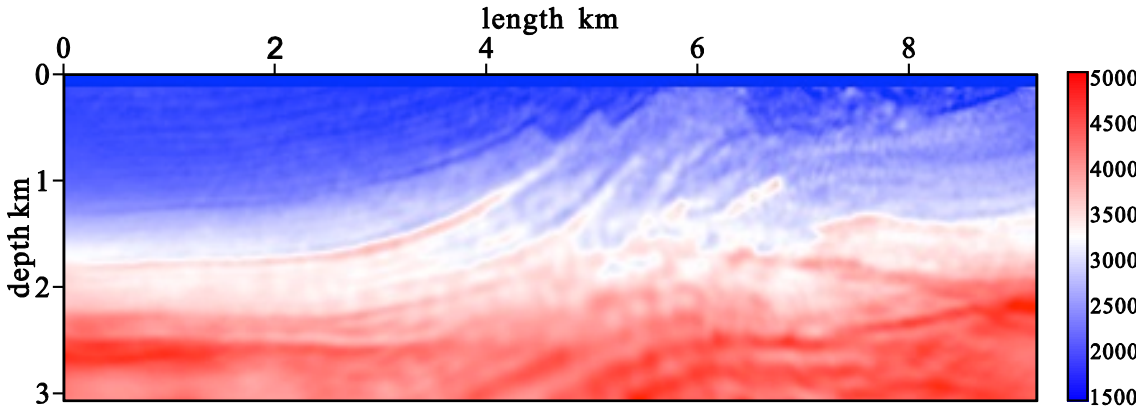
3. $\delta$ is just a constrain parameter to avoid the very large and small update. It needs to be larger than zero. Usually we could choose a suitable $\delta$ based on the update value. In this test we choose $\delta = 10^{-4}$.

4. The choice of $M$ plays an important role for the success of NLBFGS based FWI. In this section, we discuss the influence of $M$ on the inversion. During the test we use fixed values for the other parameters ($\alpha = 2$, $\rho = 0.2$, $\delta = 10^{-4}$).

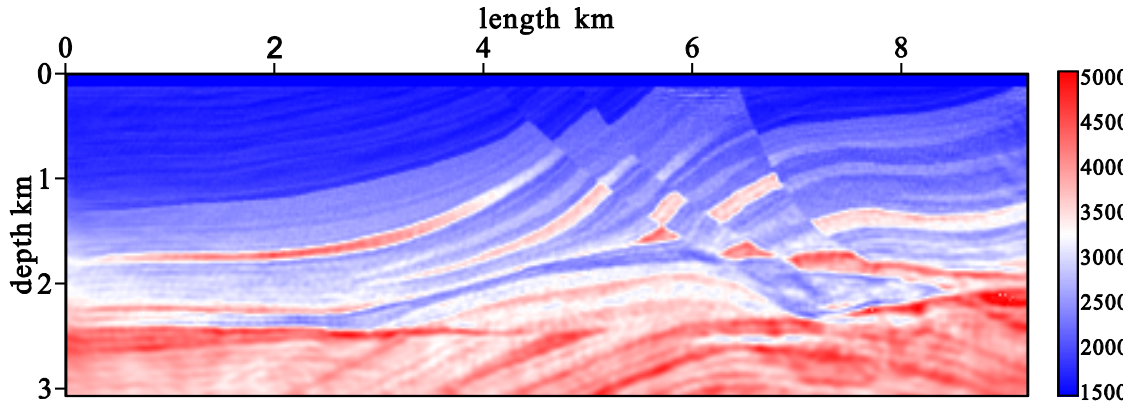
Table 3 illustrates some information of the FWI method with different $M$. Fig. 6 shows the inversion results for different $M$. When $M = 0$, we could find $m(k) = 0$ and $j = 0$, the equation

$$f(x_k + \alpha_k d_k) \le \max_{0 \le j \le m(k)} \{f(x_{k-j})\} + \delta\alpha_k d_k^T g_k$$

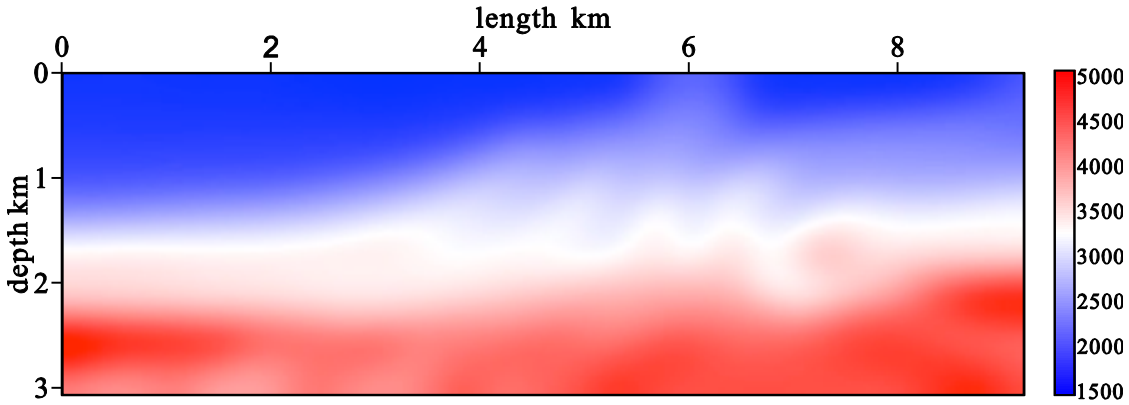
could be written as $f(x_k + \alpha_k d_k) \le f(x_k) + \delta\alpha_k d_k^T g_k$. In this case we will change NLBFGS based FWI to LBFGS based FWI. From Table 3 we could find that as the value of *M* increased, the iteration time is reduced and the inversion precision is also increased. But when *M* is too large (*M*=6), it may not be suitable for the linear search condition. According to our test, *M* = 2 is a good start value.




(a) Inversion result with ρ = 0.01

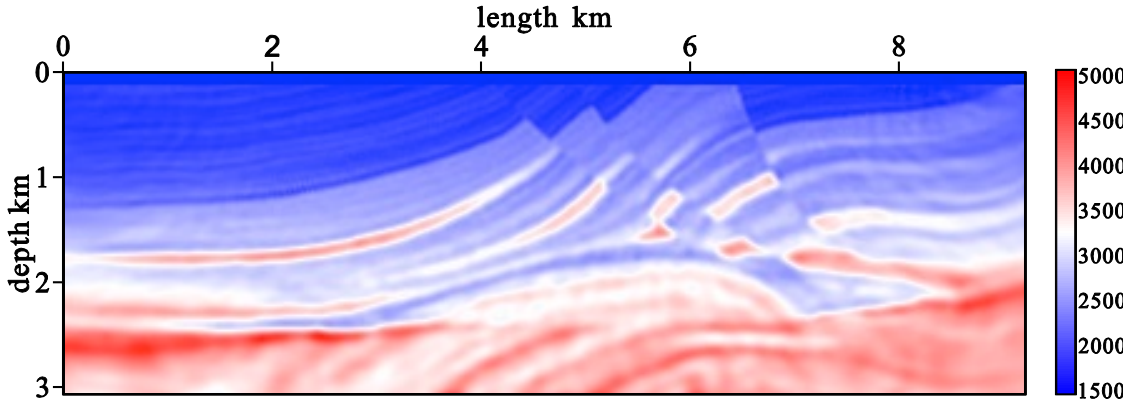



(b) Inversion result with ρ = 0.02

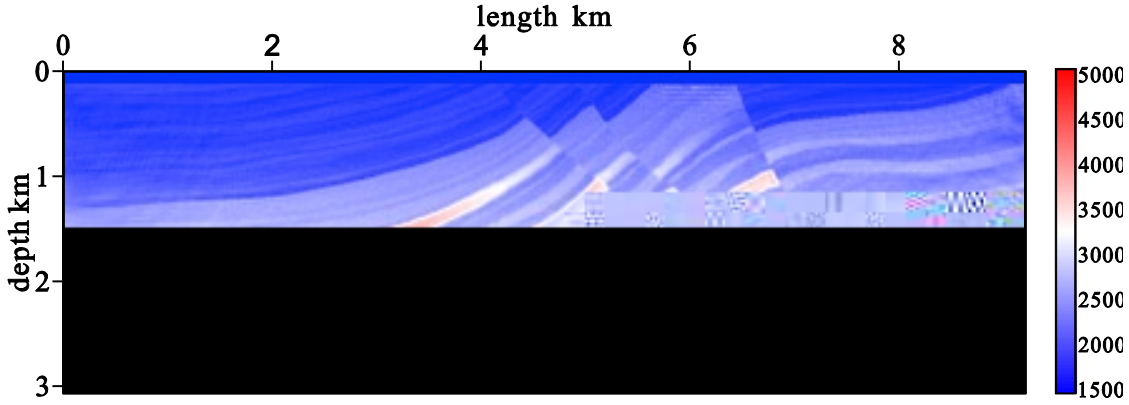



(c) Inversion result with ρ = 0.99

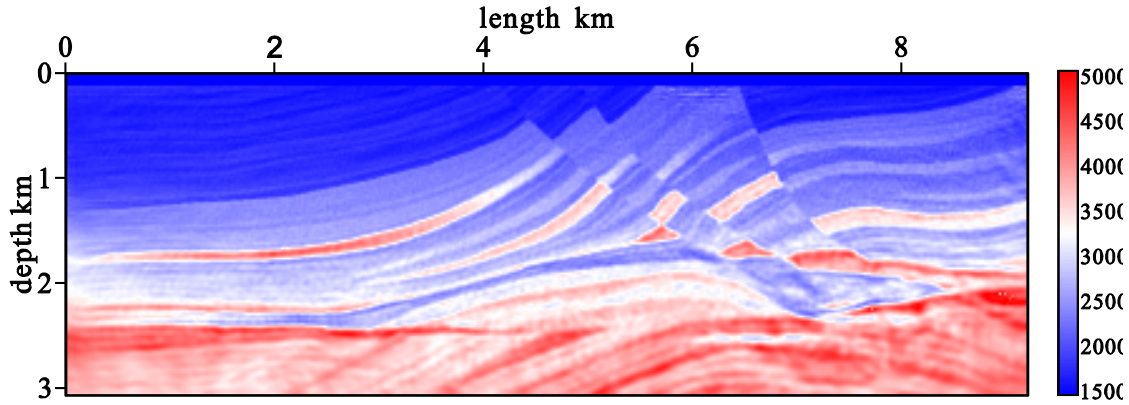
Fig. 5. Inversion results with (a) ρ = 0.01, (b) ρ = 0.2, and (c) ρ = 0.99.




(a) Inversion result with M=0




(b) Inversion result with M=1




(c) Inversion result with M=2

Fig. 6. Inversion results with M=0(a), M=1(b) and M=2(c).

Table 3. The influence of M on the inversion results.

| $M$ （$\alpha = 2$，$\rho = 0.2$） | Time(s) | Total iteration number | $E_{res}$(%) |
|---|---|---|---|
| $M$=0 | 123079 | 408 | 11.98 |
| $M$=1 | 122878 | 326 | 10.83 |
| $M$=2 | 119578 | 285 | 10.52 |
| $M$=6 | 236762 | 2500 | ------- |

As shown in Figs. 6(a) and 6(c), these represents the FWI results for the LBFGS and NLBFGS method. Comparing them with the true model, we can find the main structures of the model are well reconstructed for both methods. The NLBFGS method gives better result compared with the LBFGS one. It gives better structure reconstruction in both shallow and deep parts of the model. This is because it avoids the local minimum point by using the non-monotone line search technique.

Fig. 7 is the comparison between the true model and the results of the FWI based on the LBFGS and NLBFGS method at receiver number 205. From Fig. 7 we could find that at the depth less than 3 km both methods give similar results compared with true model. However, the NLBFGS method gives slightly better results when the depth is more than 3 km.

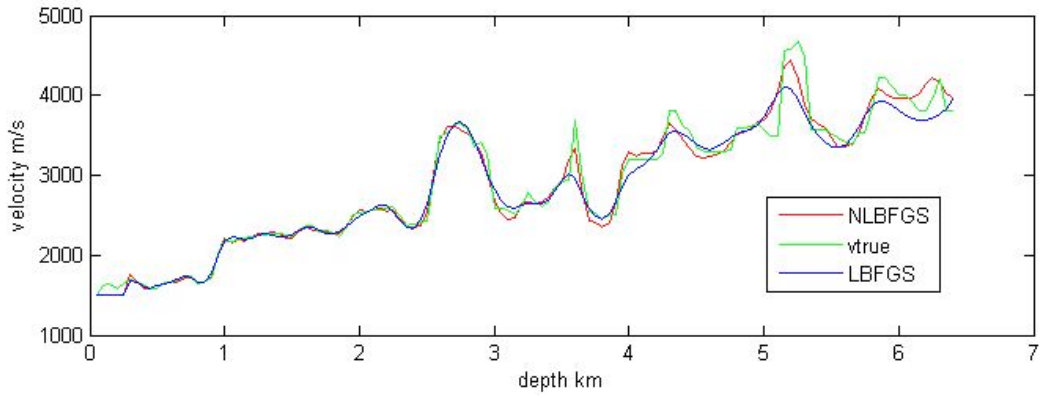



Fig. 7. Velocity comparison at receiver number 205.

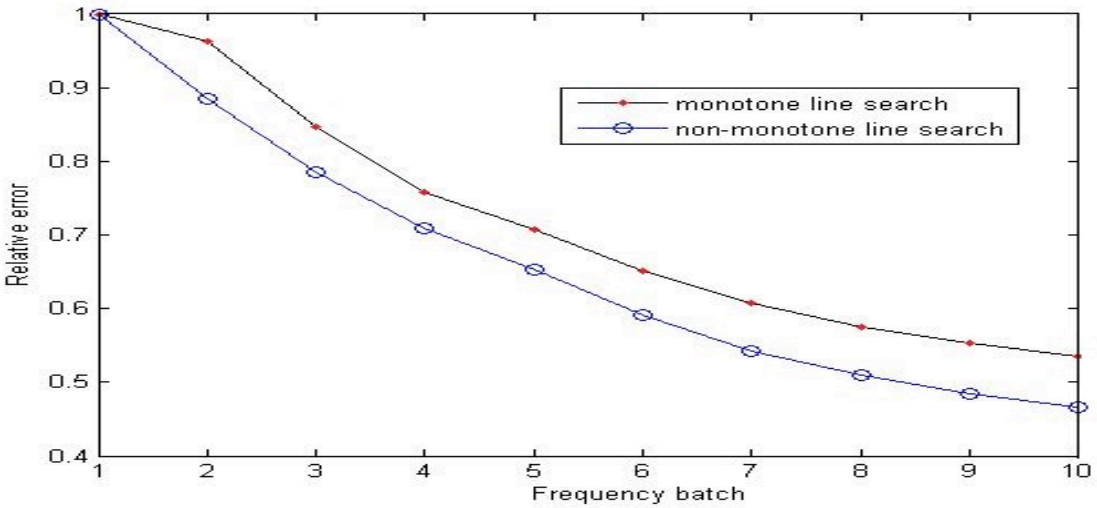



Fig. 8. Relative model error vs. frequency batch.

Fig. 8 shows the relative model errors plotting with frequency. This figure gives a comparison between the results based on the LBFGS ($M$=0, red dots) and NLBFGS ($M$=2, blue circles) method. Comparing with the monotone line search, the non-monotone line search has a better convergence rate.

## STREAMER DATA EXAMPLE

We also test our FWI method on a two-dimensional conventional streamer data. The length of the towline is about 8.1 km and the recording geometry consists of 500 shots and 81000 receivers. The shot position is 100 m from the first receiver and the distance between each receiver is 50 m. The data is prepossessed by using standard marine data processing steps. The key processing steps include low pass filtering, amplitude compensation and de-multiple. We also apply a 3D to 2D correction to the data sets which is necessary processing step for the FWI (Wang, 2009). As the amplitude between the synthetic and real data set is quite different. We apply an amplitude correction to normalize both the synthetic and real data amplitude to 1. This means we basically do the phase only FWI here. Fig. 9 shows an example processed shot gather with different frequency filter. Based on that we determine the initial inversion frequency is 3.5 Hz. The inversion frequency range is from 3.5 to 15 Hz with 0.6 Hz interval. After several tests, we decide to not use frequency group and invert each frequency separately. In total, 19 frequencies are used and 60 iterations are done for each frequency. The inversion model size is 376×2327 and the grid size is 25 m. For the initial parameters for

the NLBFGS FWI we use $\alpha = 2$， $\rho = 0.2$， $\delta = 10^{-4}$, $M = 2$.

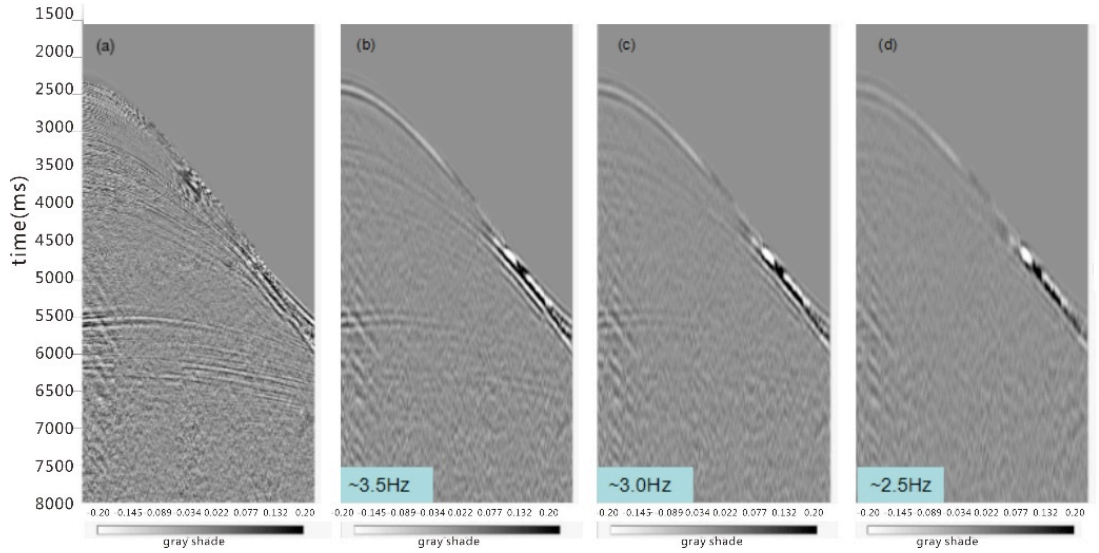



Fig. 9. Example shot gather with different frequency range. (a) Processed shot gather, (b) low pass filter to 3.5 Hz, (c) low pass filter to 3 Hz, (d) low pass filter to 2.5 Hz.

The starting model is also key aspect for the success of the FWI problem since we only use the local optimization algorithm. To avoid cycle-skipping problem we need a relative good starting model. In this paper, we use a starting velocity model got by combine the migration velocity and the traveltime tomography as shown in Fig. 10. Fig. 11 shows the synthetic shot gather generated by the starting model. We could find that it somehow reproduce the first several cycles of the real shot gather which gives us some confidence of the starting model.

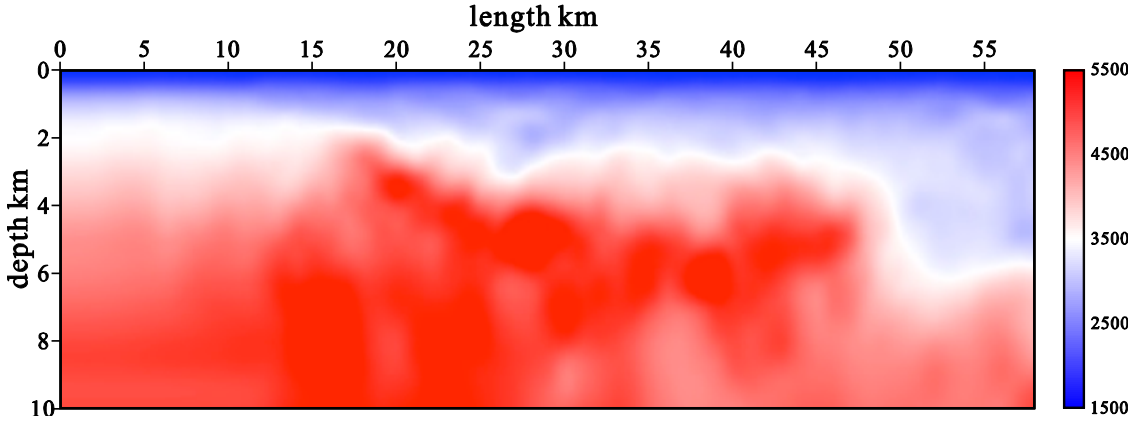



Fig. 10. The starting model.

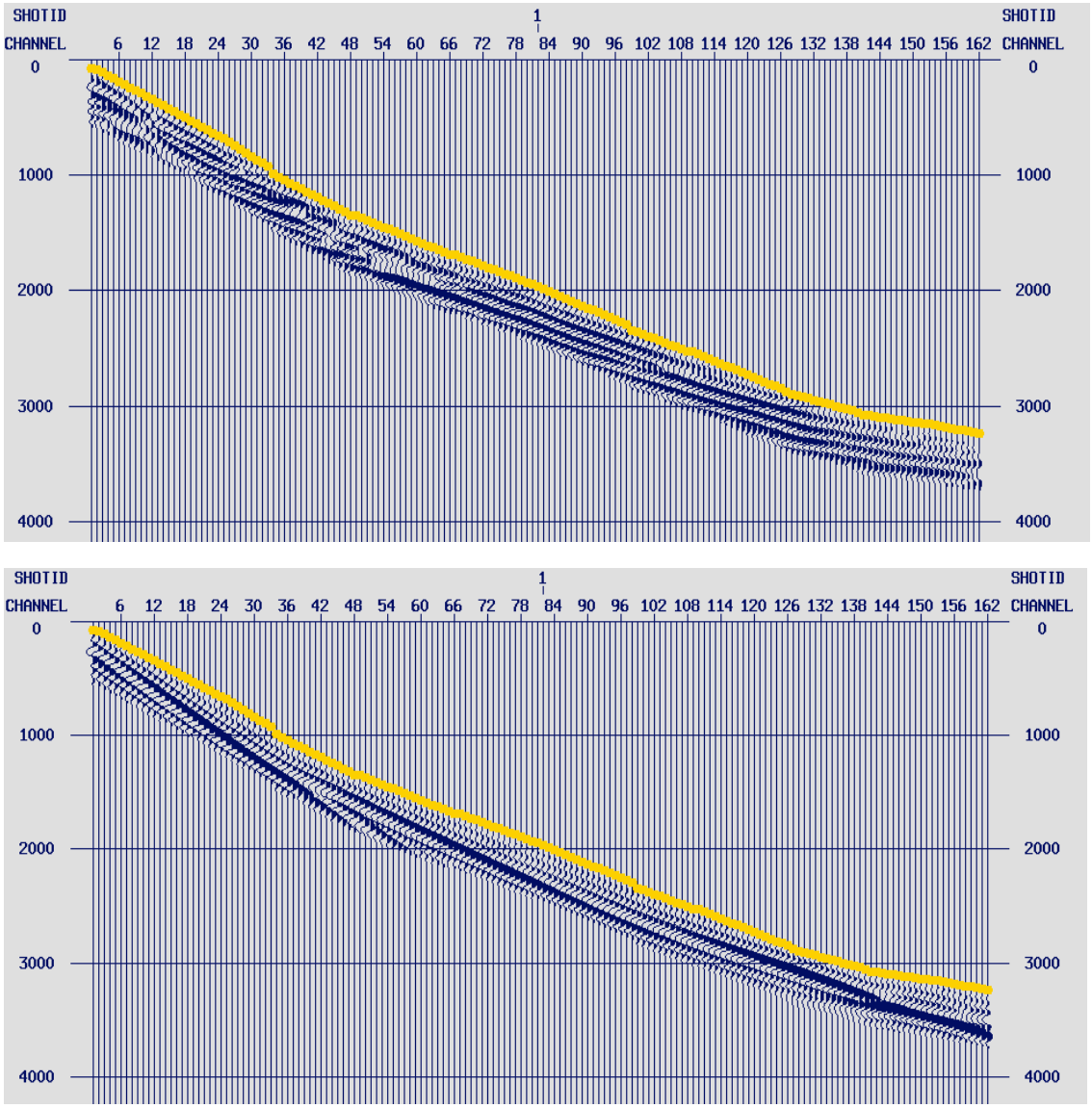
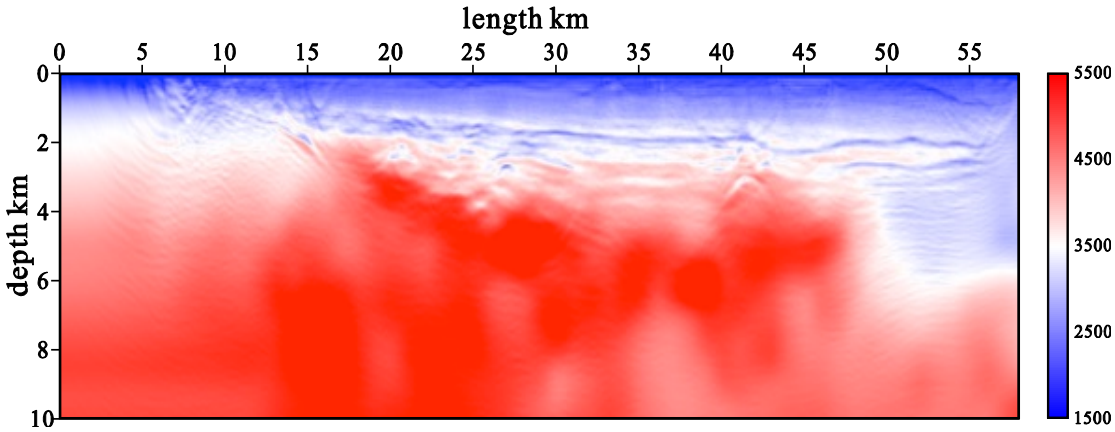
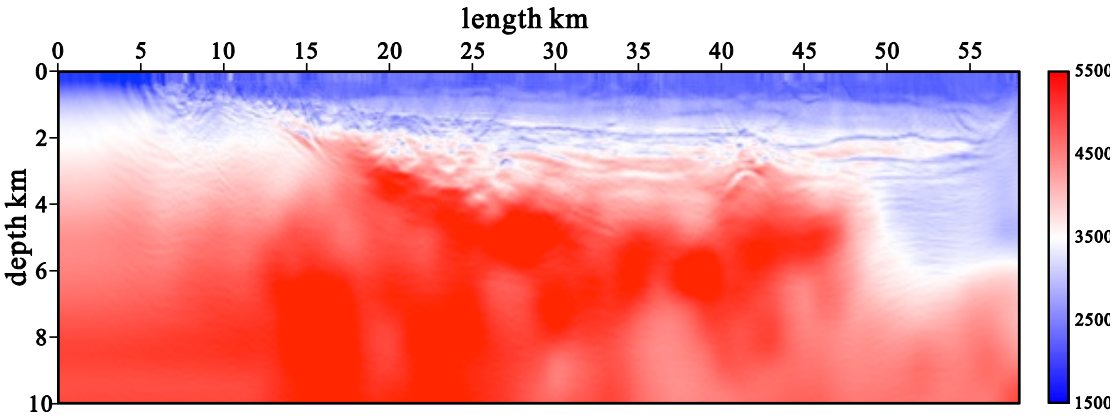



Fig. 11. Shot gather comparison between real data (upper) and synthetic data generated by the starting model (bottom).

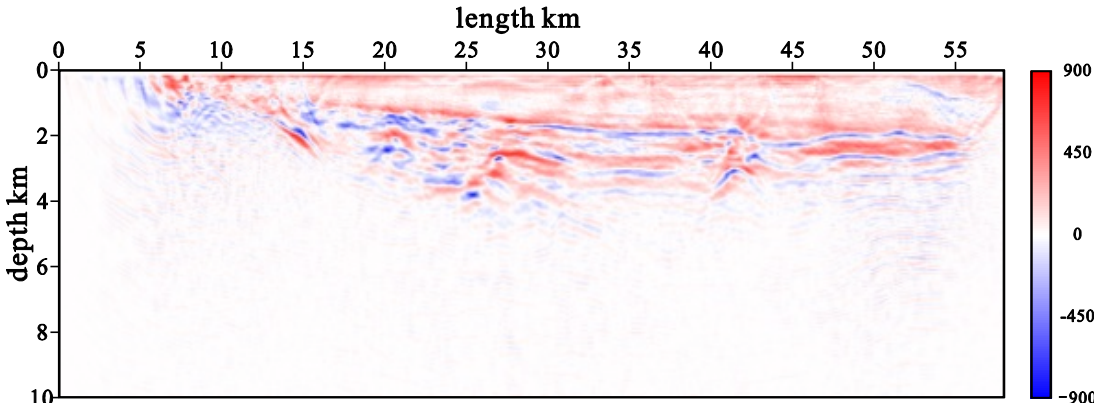



(a) The result of the LBFGS based FWI

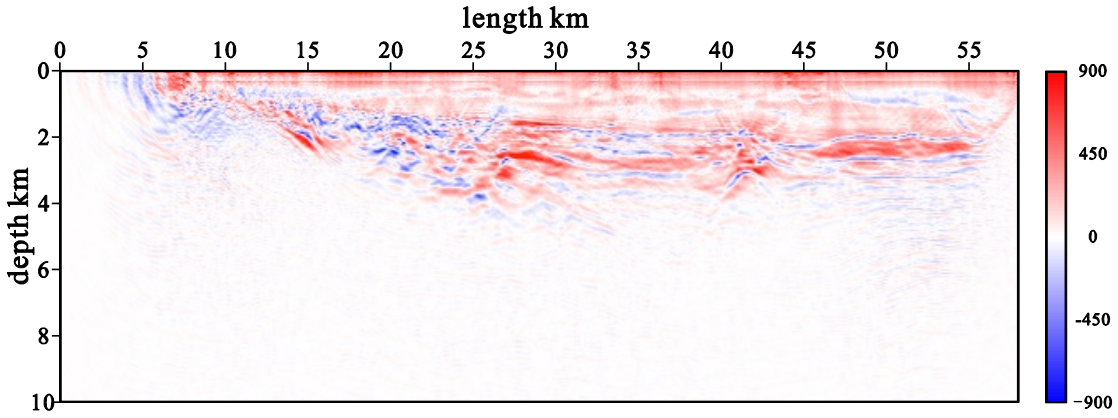



(b) The result of the NLBFGS based FWI

Fig. 12. The results of the LBFGS based and NLBFGS based FWI.




(a) The velocity update for the LBFGS based FWI




(b) The velocity update for the NLBFGS based FWI

Fig. 13. The velocity updates of the LBFGS based and NLBFGS based FWI.

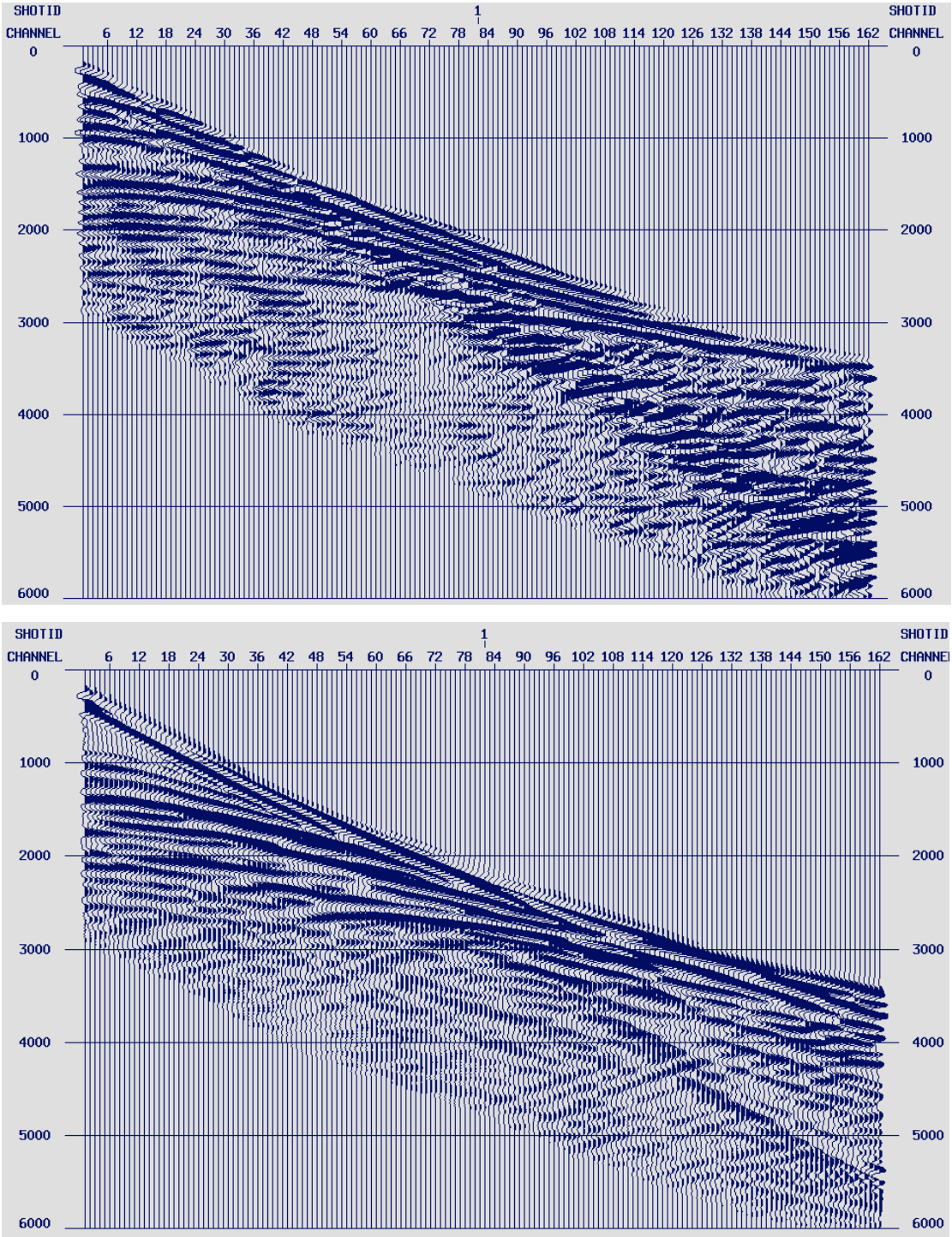
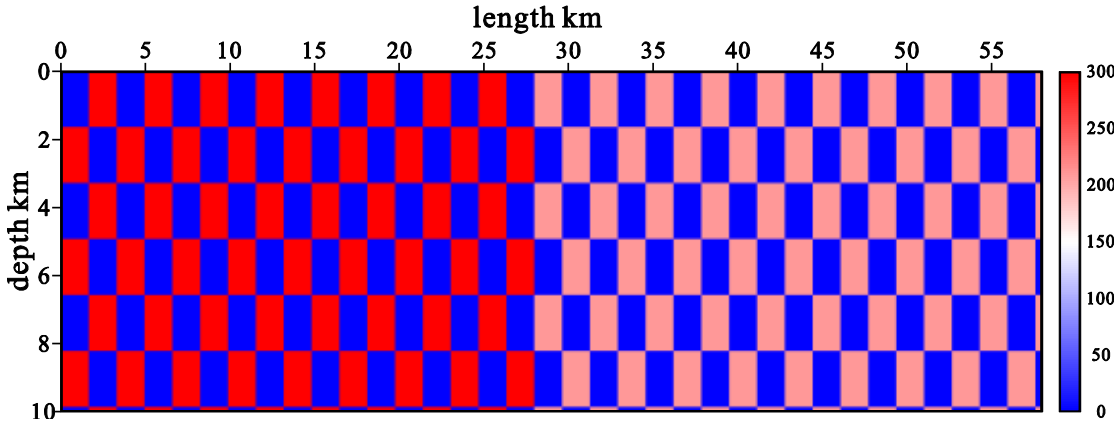
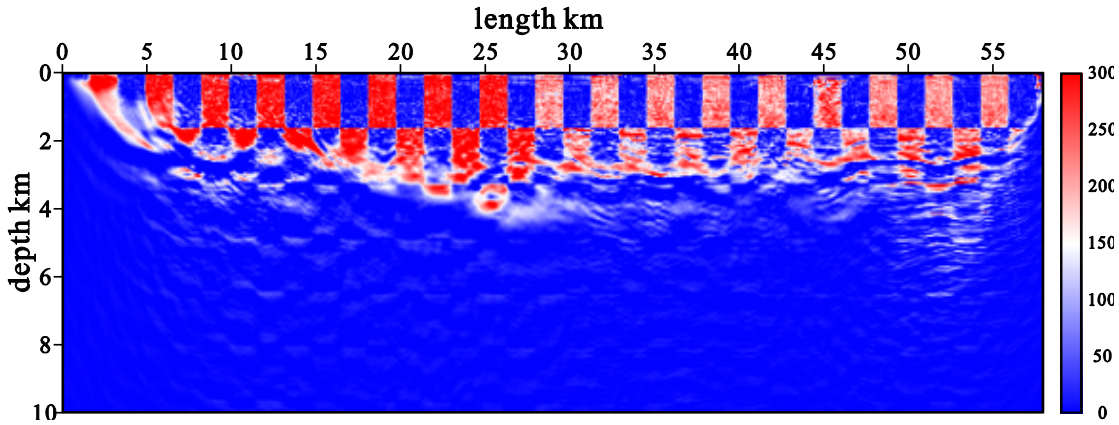



Fig. 14. Shot gather comparison between real data (upper) and synthetic data generated by the NLBFGS FWI final model (bottom).

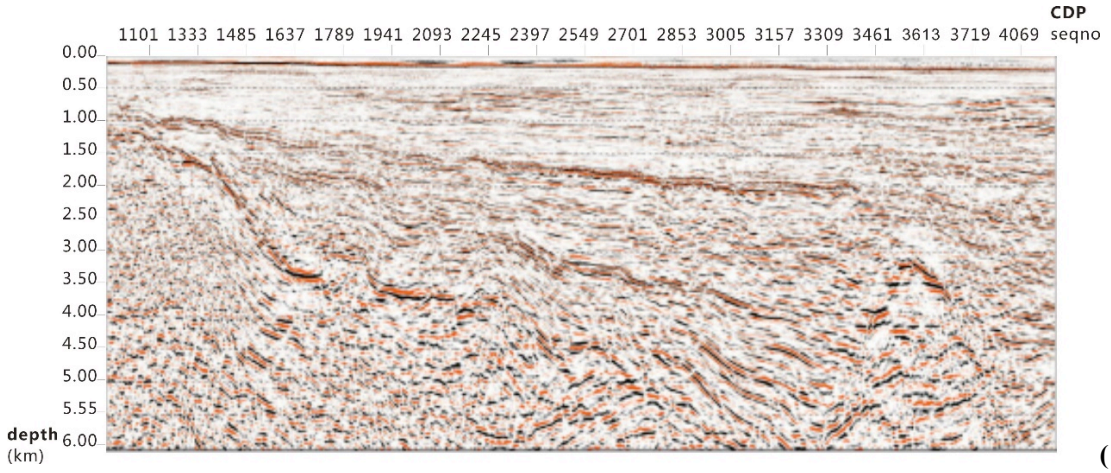



(a) True checker pattern for the checker board test.

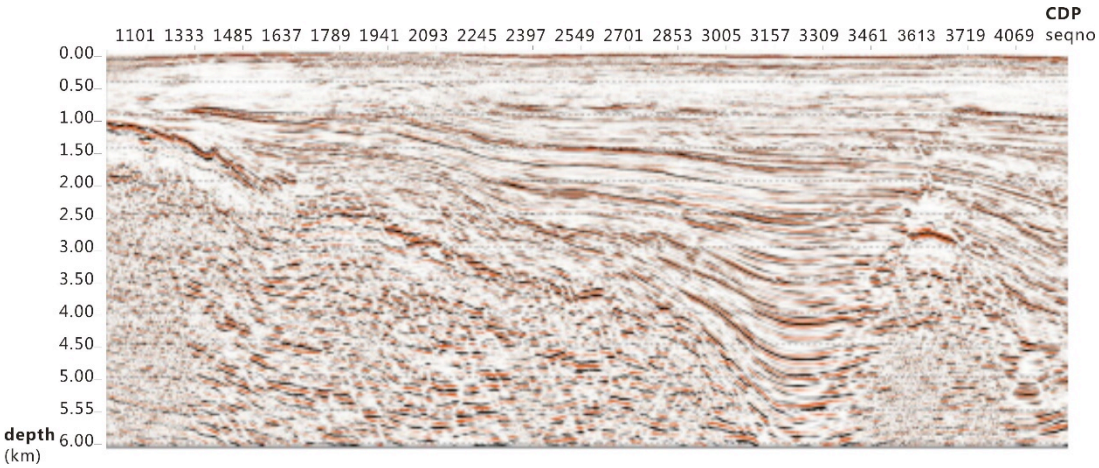



(b) Recovered checker pattern.

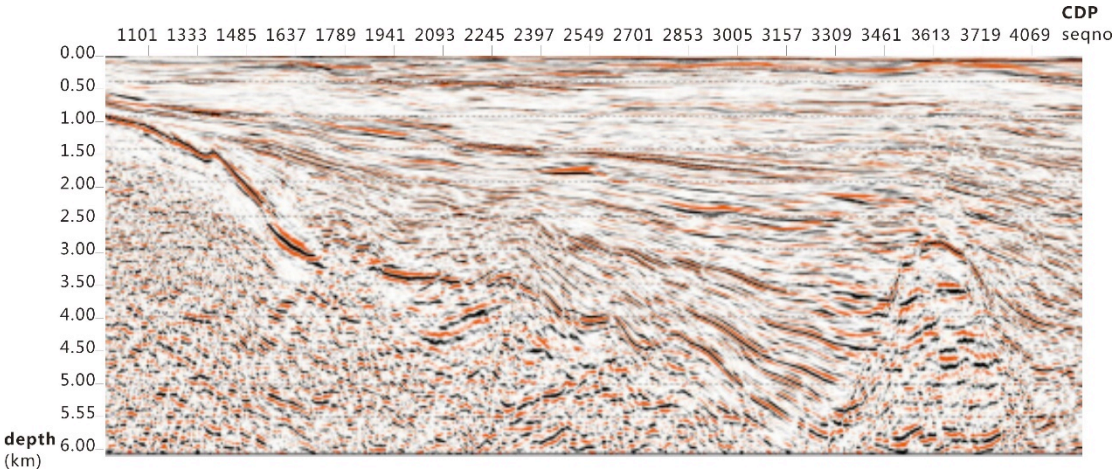
Fig. 15. Checker board test for the NLBFGS FWI.




(a) The migrated section based on the starting model




(b) The migrated section based on the LBFGS based FWI result model




(c) The migrated section based on the NLBFGS FWI result model

Fig. 16. The Kirchhoff pre-stack depth migration results based on (a) the starting model, (b) the LBFGS based FWI and (c) the NLBFGS FWI.

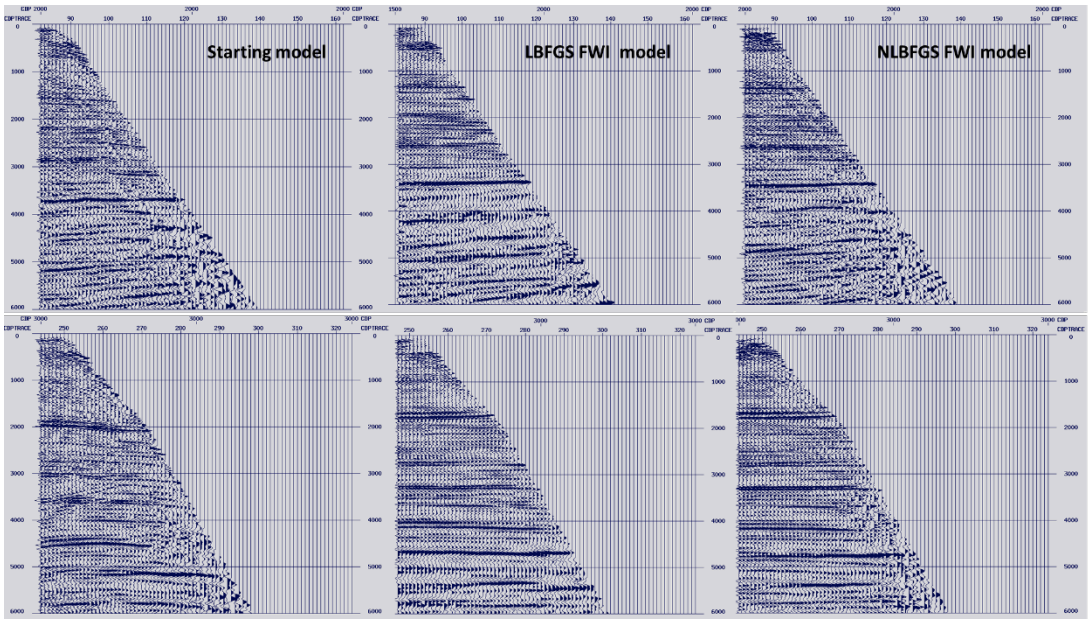



Fig. 17. Two common image gather examples from different parts (upper and low figure) of the profile using different velocity model.

Fig. 12 shows the comparison of the FWI results between the LBFGS and the NLBFGS method. Fig. 13 shows the velocity updates of both inversion method. Both methods improve the model resolution and the main improvement is located in the shallow parts (from 0 to 3 km). We could also find that the NLBFGS method gives some different structure updates compared with the conventional LBFGS method. Fig. 14 shows the shot gather comparison between the real shot gather and synthetic shot gather generated by the final result model with the NLBFGS method. The synthetic shot gather shows some similarities with the real shot gather in early arrival waveforms. Although it is not a perfect match, this provides us a quality control of the final results. We also run a checker board test to check the resolution of the NLBFGS method. This is carried out by perturbing the velocity model by +2-5% with a 1650 by 1650 m checker. The checker pattern and recovered checker pattern is shown in Fig. 15. We could clearly see that the checker is well recovered in the shallow parts of the model (from 0 to 3 km). This gives us some confidences of our final velocity model at least in the shallow parts. Fig.16 shows the Kirchhoff pre-stack depth migration results of the starting model, the conventional LBFGS based FWI method and the NLBFGS based FWI method. From the results we could find the NLBFGS result model gives best migration image. It improves the migration image and removes some artifacts structures. Fig. 17 shows two common image gather examples from different parts of the profile using different velocity model. Some reflections are more coherent and flatter in the gathers using NLBFGS FWI velocity model compared with the one using LBFGS FWI model. This is consistent with the differences between the migration sections obtained using the LBFGS and NLGFGS FWI velocity models.

## CONCLUSIONS

The non-monotone line search technique could increase the convergence speed of FWI and avoid the local minimum point. The LBFGS method is generally regarded as the most effective quasi Newton method for optimization problems because of its stability, convergence and limited internal storage use. In this paper, we combine the non-monotone line search technique and the LBFGS method into the FWI method. We test the new method with both synthetic data and real streamer data. The synthetic example shows that the new method could provide higher accuracy and faster converge rate of the inversion. We also discuss the influence of initial parameters on the inversion results for our method. Try to give some general guideline for this method. The streamer data example shows the feasibility of applying the NLBFGS FWI method on the real data. The new method gives similar results compared with the conventional method but with less artifacts. It also improves the migration image in the shallow and even some deeper parts.

## ACKNOWLEDGMENTS

We want to thank the National Natural Science Foundation of China (Grant No. 41404092 and No. 41304086) for their support of this work. CLOBE Claritas$^{TM}$ under license from the Institute of Geological and Nuclear Sciences Limited, Lower Hutt, New Zealand, was used to plot some seismic gathers.

## REFERENCES

Abdi, H. and Williams, L., 2010. Principal component analysis: Wiley Interdisciplinary Reviews. Computat. Statist., 2: 433–459.

Benhadjali, H., Operto, S. and Virieux, J., 2011. An efficient frequency-domain full waveform inversion method using simultaneous encoded sources, Geophysics, 76(4): 109-124. DOI:10.1190/1.3581357.

Brossier, R., 2011. Two-dimensional frequency-domain visco-elastic full waveform inversion: Parallel algorithms, optimization and performance. Comput. Geosci., 37: 444-455. DOI:10.1016/j.cageo.2010.09.013.

Broyden, C.G., 1965. A class of methods for solving nonlinear simultaneous equations. Mathemat. Computat., 19(92): 577-593. DOI:10.1090/S0025-5718-1965-0198670-6.

Broyden, C.G., 1967. Quasi-Newton methods and their application to function minimization. Mathemat. Computat., 21(99): 368-381. DOI:10.2307/2003239.

Byrd, R.H. and Nocedal, J., 1989. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. SIAM J. Numer. Analys., 26: 727-739. DOI:10.1137/0726042.

Cheng, W.Y. and Li, D.H., 2010. Spectral scaling BFGS method. J. Optimizat. Theory Applicat., 146: 305-319. DOI:10.1007/s10957-010-9652-y.

Grippo, L., Lampariello, F. and Lucidi, S., 1986. A nonmonotone line search technique for Newton's Method. SIAM J. Numer. Analys., 23: 707-716. DOI:10.1137/0723046.

Grippo, L., Lampariello, F. and Lucidi, S., 1989. A truncated Newton method with nonmonotone line search for unconstrained optimization. J. Optimizat. Theory Applicat., 60: 401-419. DOI:10.1007/BF00940345.

Han, J. and Liu, G., 1997. Global convergence analysis of a new nonmonotone BFGS algorithm on convex objective functions. Computat. Optimizat. Applicat., 7: 277-289. DOI:10.1023/A:1008656711925.

Huang, Z.H., Hu, S.L. and Han, J.Y., 2009. Convergence of a smoothing algorithm for symmetric cone complementarity problems with a non-monotone line search. Sci. China, 52: 833-848. DOI:10.1007/s11425-008-0170-4.

Liu, C.C., Han, M., Han, L.G., Huang, F. and Deng, W., 2012. Application of principal component analysis for frequency-domain full waveform inversion. Expanded Abstr., 82nd Ann. Internat. SEG Mtg., Las Vegas: 1-5. DOI:10.1190/segam2012-0909.1.

Liu, D.C. and Nocedal, J., 1989. On the Limited Memory BFGS Method for large scale optimization. Mathemat. Program., 45: 503-528. DOI:10.1007/BF01589116.

Ma, Y.and Hale, D., 2012. Quasi Newton full-waveform inversion with a projected Hessian matrix. Geophysics, 77(5): 207-216. DOI:10.1190/geo2011-0519.1.

Moore, B., 1981. Principal component analysis in linear systems: Controllability, observability, and model reduction. IEEE Transact. Automat. Contr., 26: 17–32.

Nocedal, J., 1980. Updating quasi-Newton matrices with limited storage. Mathemat. Comput., 35: 773-782. DOI: 10.2307/2006193.

Panier, E.R. and Tits, A.L., 1991. Avoiding the Maratos effect by means of a non-monotone linear search I , general constrained problems. SIAM J. Numer. Analys., 28**:** 1183-1195.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. Geophysics, 64: 888-901. DOI:10.1190/1.1444597.

Raydan, M., 1997. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM J. Optimizat., 7: 26-33. DOI:10.1137/S1052623494266365.

Schiemenz, A. and Igel, H., 2013. Accelerated 3-D full-waveform inversion using simultaneously encoded sources in the time domain: application to Valhall ocean-bottom cable data. Geophys. J. Internat., 195: 1970-1988. DOI:10.1093/gji/ggt362.

Shi, Z.J. and Shen, J., 2006. Convergence of non-monotone line search method. J. Computat. Appl. Mathemat., 193: 397-412. DOI:10.1016/j.cam. 2005. 06.003.

Sun, W.Y., Han, J.Y., Sun, J., 2002. Global convergence of non-monotone descent method for unconstrained optimization problems. J. Computat. Appl. Mathemat., 146: 89-98. DOI:10.1016/S0377-0427(02)00420-X.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation. Geophysics, 49: 1259-1266. DOI:10.1190/1.1441754.

Wang, Y.H. and Rao, Y., 2009. Reflection seismic waveform tomography. J. Geophys. Res., 114(B3): 63-73. DOI:10.1029/2008JB005916

Zhang, H.C. and Hager, W.W., 2004. A Non-monotone Line search technique and its application to unconstrained optimization. SIAM J. Optimizat., 14: l043-1056. DOI:10.1137/S1052623403428208.

Zhao, H. and Gao, Z.Y., 2005. Equilibrium algorithms with non-monotone line search technique for solving the traffic assignment problems. J. Systems Sci. Complex., 18: 543-555.

Zhu, D.T., 2003. Non-monotonic back-tracking trust region interior point algorithm for linear constrained optimization. J. Computat. Appl. Mathemat., 155: 285-305. DOI:10.1016/S0377-0427(02)00870-1.