Memory-efficient frequency-domain Gauss–Newton method for wave-equation first-arrival traveltime inversion

Jianhua Wang1 and Liangguo Dong1,2*

1School of Ocean and Earth Science, Tongji University, Shanghai, 200092, China, and 2State Key Laboratory of Marine Geology, Tongji University, Shanghai, 200092, China

Received October 2020, revision accepted April 2021

ABSTRACT

Wave-equation traveltime inversion (WTI) can be used to automatically obtain a background near-surface velocity model (NSM), which overcomes the high-frequency approximation in ray theory. It is generally implemented in the time domain. However, the commonly used gradient-based optimisation methods (such as the steepest-descent method) in WTI have a low convergence rate and may yield less accurate results within limited iterations in geologically complex regions. To increase the convergence rate and improve the inversion accuracy, we propose a frequency-domain truncated Gauss–Newton first-arrival wave-equation traveltime inversion (GN-WTI) method to retrieve the background NSM. As only a few frequencies are used for inversion, the proposed frequency-domain WTI method significantly reduces the computational memory requirements by more than two orders of magnitude in comparison with the conventional time-domain WTI method. Therefore, the proposed method is especially advantageous for the building of large three-dimensional models. In this GN-WTI method, according to the derived explicit traveltime residual kernel, the gradient and Hessian vector products can be computed efficiently using an elegant and improved scattering integral approach as long as the source-side wavefields and non-redundant receiver-side Green’s functions are computed and stored in advance. The conjugate gradient approach is used to solve the Gauss–Newton normal equation to obtain the Gauss–Newton direction in inner loops. Here, the Gauss–Newton Hessians of the ray-based traveltime inversion and WTI are compared to demonstrate the advantages of WTI. The trial runs with a simple periodic velocity model example showed that the proposed GN-WTI method outperforms the WTI method when using the steepest-descent and limited-memory Broyden–Fletcher–Goldfarb–Shanno approaches in terms of the convergence rate and inversion accuracy. A complex Marmousi model was further used to illustrate the effectiveness of GN-WTI. The proposed method should be beneficial in near-surface velocity model building.

Key words: Near-surface, Seismic, Traveltime, Velocity.

INTRODUCTION

The determination of the background near-surface velocity model (NSM) is crucial for imaging the Earth’s deeper structures at various scales, in application ranging from oil/gas seismic exploration to crust–mantle structure studies (Zhang and Toksöz, 1998; Tape et al., 2010; Zhao et al., 2012; Cova et al., 2018). In oil/gas seismic exploration, the NSM is indispensable to compensate for the effects of variations in the low-velocity weathering zone and elevation (Sheriff, 2002).
Furthermore, an accurate NSM is essential for the successful imaging of deeper structures. An initial NSM containing accurate low-to-intermediate-wavenumber velocity information also plays an important role in the prevention of the phenomenon in which the full waveform inversion (FWI) can become trapped in local minima (Virieux and Operto, 2009). Owing to the quasi-linear property of seismic traveltime with respect to velocity, the first-arrival traveltimes are frequently used to invert the background NSM. Ray-based algorithms have been used widely in the oil exploration industry for decades to obtain shallow velocity structures (Olson, 1989; Zhu et al., 1992; Sei and Symes, 1994; Zelt et al., 2006). However, for geologically complex regions, inaccurate inversion results may be produced in response to the high-frequency approximation in ray theory (Wang et al., 2008). Moreover, ray-based methods rely heavily on correct first-arrival pickings. Although traveltime picking can be automated, it remains time-consuming because manual interventions are often required for quality control (Keho and Zhu, 2009).

To consider the finite-frequency nature of seismic propagation during traveltime inversion, Luo and Schuster (1991) developed a wave-equation traveltime inversion (WTI) method. Compared with the conventional ray-based traveltime inversion (RTI), WTI has several advantages in that it offers an automatic wave-equation-based tomography method suitable for geologically complex regions and free of traveltime picking. In previous studies (Luo and Schuster, 1991; Zhou et al., 1995; Van Leeuwen and Mulder, 2010; Ma and Hale, 2013; Luo et al., 2016), WTI has been solved by gradient-based approaches (such as the steepest-descent approach) in the time domain. This involves a forward and backward propagation process and temporal zero-lag cross-correlation of the forward and adjoint wavefields. However, the steepest-descent method has a low convergence rate and may yield less accurate results within limited iterations in geologically complex regions. In contrast, the Gauss–Newton optimization can achieve near-quadratic convergence in the neighbourhood of the minimum, and, thus, it outperforms the linear convergence of the steepest-descent optimization. Moreover, the Hessian matrix can scale and filter the gradient, which leads to a more accurate inversion result. Therefore, the second-order derivation information of the misfit function provided by the Hessian matrix is very helpful in WTI. Zheng et al. (2013) compared the steepest-descent, conjugate gradient (CG) and limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) optimization methods in WTI. These results showed that L-BFGS provides a higher convergence rate and can be used to obtain a velocity model with higher accuracy compared with the other two gradient methods. In the L-BFGS algorithm, the gradient vector is implicitly preconditioned with an approximation matrix of inverse Hessian. The approximation in the L-BFGS algorithm depends on the gradients of the previous iterations. However, the local curvature information estimated by previous iterations can be inaccurate.

The Gauss–Newton approach represents a more accurate optimization method for utilizing Hessian information. In this study, we propose a truncated Gauss–Newton first-arrival wave-equation traveltime inversion (GN-WTI) method implemented in the frequency domain to invert the NSM. The proposed frequency-domain GN-WTI method can retrieve the background NSM using limited frequencies with sparse frequency samples. Therefore, the computational memory requirements are greatly reduced compared with the conventional time-domain WTI method. In the GN-WTI method, the gradient vector and Hessian vector products are computed efficiently by the improved scattering integral (SI) approach, which involves only simple matrix vector multiplication and addition operations, as long as the source-side wavefields and non-redundant receiver-side Green’s functions are computed and stored in memory beforehand. A numerical example with a simple periodic velocity model demonstrated that the proposed method outperforms the WTI methods using the steepest-descent and L-BFGS approaches in terms of the convergence rate and inversion accuracy. Then a complex Marmousi model was applied to further illustrate the effectiveness of our method.

**METHODOLOGY**

First-arrival traveltime inversion can be formulated as a least-squares problem, in which the L2 norm based misfit function $J$ is usually used to evaluate the difference between the observed and calculated first-arrival traveltimes:

$$J = \frac{1}{2} \| \Delta t \|^2.$$  \hspace{1cm} (1)

where $\Delta t$ is the traveltime residual vector, and it is defined as equation (2):

$$\Delta t = (\Delta t_1, \Delta t_2, \ldots, \Delta t_m)^T.$$  \hspace{1cm} (2)

Here, $T$ denotes the transpose operator. $\Delta t_i$, $i = 1, 2, \ldots, m$, is the individual traveltime residual corresponding to the $i$th source–receiver pair, which can be estimated by cross-correlating the observed and calculated first arrivals. $m$ is the number of total source–receiver pairs. This
least-squares problem is usually solved by iteratively updating the model, as demonstrated in equation (3):

$$v_{k+1} = v_k + \gamma_k p^\text{GN}_k.$$  \hspace{1cm} (3)

Here, $\gamma_k$ and $p^\text{GN}_k$ denote the step-length and the Gauss–Newton direction at the $k$th iteration, respectively. In each iteration, the Gauss–Newton direction can be obtained by solving the Gauss–Newton equation:

$$H_p p^\text{GN} = -\nabla v_J. \hspace{1cm} (4)$$

where $\nabla_J$ denotes the gradient vector of the misfit function and $H_p$ is the Gauss–Newton Hessian matrix. The gradient and Gauss–Newton Hessian of the misfit function $J$ can then be expressed as equation (5):

$$\nabla_J = \sum_{i=1}^m \frac{\partial \Delta t_i}{\partial v_i} \Delta t_i = K^T \Delta t, \hspace{1cm} (5)$$

and equation (6):

$$H_p = K^T K. \hspace{1cm} (6)$$

Here, $\frac{\partial \Delta t_i}{\partial v_i}$ is the sensitivity kernel of the individual traveltime residual with respect to velocity. This constitutes the $m \times n$ Jacobian matrix $K$:

$$K = \begin{bmatrix} \frac{\partial \Delta t_1}{\partial v_1}^T \\ \frac{\partial \Delta t_2}{\partial v_1}^T \\ \vdots \\ \frac{\partial \Delta t_m}{\partial v_1}^T \end{bmatrix} \hspace{1cm} (7)$$

where $n$ is the number of model spatial grids.

In previous studies (Luo and Schuster, 1991; Zhou et al., 1995; Van Leeuwen and Mulder, 2010; Ma and Hale, 2013; Luo et al., 2016), the negative gradient of the misfit function was usually taken as the search direction in WTI. However, this does not fully take into account the scale and filter effect of the Hessian. As stated in the previous section, the steepest-descent approach in the time domain has a low convergence rate and may yield less accurate inversion results within limited iterations in geologically complex regions. Here, we used the Gauss–Newton optimization method in the frequency domain to implement the first-arrival WTI. The reason for not using the exact Hessian was that the influence of the second term in the exact Hessian is generally limited unless the doubly scattered waves are very strong, which only occurs in rare cases (Liu et al., 2020). The Gauss–Newton search direction $p^\text{GN}$ can be obtained by solving the Gauss–Newton normal equation (4) using the CG approach.

### Improved scattering integral method

The improved SI method was used to calculate the gradient (equation (5)) and Hessian vector products.

The gradient is computed by the product of the Jacobian matrix transpose, $K^T$, with the traveltime residual vector. Using the improved SI method, the gradient can be computed as

$$\frac{\partial J}{\partial v(x)} = K^T \Delta t = \begin{bmatrix} k_{11} \cdots k_{1i} \cdots k_{1m} \\ \vdots \cdots \cdots \vdots \\ k_{n1} \cdots k_{ni} \cdots k_{nm} \end{bmatrix} \begin{bmatrix} \Delta t_1 \\ \vdots \\ \Delta t_n \end{bmatrix} \hspace{1cm} (8)$$

Note that every column of the matrix $K^T$ represents an individual traveltime residual kernel corresponding to a source–receiver pair. The gradient can be obtained by accumulating the products of the individual traveltime residual kernel of a source–receiver pair with the corresponding traveltime residual. The Hessian vector products in the inner CG loop can be efficiently computed by kernel vector products $Kp$ (equation (9)), followed by the transposed kernel vector products $K^T(Kp)$ (equation (10)):

$$q = Kp. \hspace{1cm} (9)$$

$$H_p = K^T q. \hspace{1cm} (10)$$

Thus, the Hessian vector products can be computed efficiently by two nested kernel vector products without additional forward modelling, as long as the source-side wavefields and non-redundant receiver-side Green’s functions are computed and stored in memory beforehand (as indicated by equations (12)–(14)). Additional details regarding the SI implementation can be found in Liu et al. (2015) and Yang et al. (2016). The detailed workflow and implementation procedure for our GN-WTI method are presented in the Appendix. Another popular technique among geophysics communities is to compute Hessian vector products with the second-order adjoint-state method (Métivier et al., 2013, 2014). However, three forward/backward simulations are required in every inner CG loop, which inevitably results in cost-prohibitive computations within a large-scale model.
Traveltime residual kernel in the frequency domain

One key issue in our method is the explicit formulation of the Jacobian element \( \partial \Delta t_i / \partial v_j \) in equation (7).

The time-domain formulation of the traveltime residual kernel can be derived by using a connective function and applying the implicit function derivation theorem (Luo and Schuster, 1991):

\[
\frac{\partial \Delta t_i}{\partial v_j} = - \int \frac{d_{\text{obs}}(x_i, x_j; t + \Delta t_i)}{\int d_{\text{obs}}(x_k, x_j; t + \Delta t_i) d_{\text{cal}}(x_j, x_k; t) \, dt} \frac{\partial d_{\text{cal}}(x_i, x_j; t)}{\partial v_i} \, dt. \tag{11}
\]

Here, \( \partial \Delta t_i / \partial v_j \) is the derivative of the \( i \)th source–receiver pair traveltime residual with respect to the \( j \)th spatial point velocity. \( d_{\text{obs}}(x_i, x_j; t) \) and \( d_{\text{cal}}(x_i, x_j; t) \) represent the observed and calculated first arrivals, respectively, which are estimated by windowing the observed and calculated common-shot gathers. \( x_i \) and \( x_j \) are the source and receiver spatial locations, respectively. \( \Delta t_i \) represent the first- and second-order time derivatives of the time-shifted observed first arrivals, respectively. Numerically calculating the traveltime residual kernel consumes large amounts of memory/disk storage space because the time-domain wavefields and Green’s functions have to be accessible simultaneously. Therefore, we apply Parseval’s theorem to equation (11) and obtain the explicit formulation of the traveltime residual kernel in the frequency domain:

\[
\frac{\partial \Delta t_i}{\partial v_j} = \int \left( \frac{\bar{i} \omega \tilde{d}_{\text{obs}}(x_i, x_j; \omega) \exp(i \omega \Delta t_i)}{E} \right)^* \frac{\partial \tilde{d}_{\text{cal}}(x_i, x_j; \omega)}{\partial v_i} \, d\omega, \tag{12}
\]

where

\[
E = \int \left( -\omega^2 \tilde{d}_{\text{obs}}(x_i, x_j; \omega) \exp(i \omega \Delta t_i) \right)^* \tilde{d}_{\text{cal}}(x_i, x_j; \omega) \, d\omega. \tag{13}
\]

The superscript \( \sim \) denotes the frequency-domain representation of the corresponding seismic data, and the symbol \( \ast \) denotes the conjugate operator. Under the constant-density acoustic approximation, the wavefield Born kernel can be expressed as follows (Woodward, 1992):

\[
\frac{\partial \tilde{d}_{\text{cal}}(x_i, x_j; \omega)}{\partial v_i} = -\frac{2 \omega^2}{\nu_i^2} \tilde{u}(x_i; \omega) G(x_i; \omega; x_j). \tag{14}
\]

Here \( \tilde{u}(x_i; \omega) \) and \( G(x_i; \omega; x_j) \) are the source-side wavefield and receiver-side Green’s function. According to equations (12)–(14), the traveltime residual kernel can be computed efficiently and the storage space requirements are significantly reduced because only a limited number of frequencies are necessary during inversion.

One key issue is how to select the frequency samples for the inversion. The principle is to select frequency samples with relatively strong energy. The frequency-domain WTI can be carried out using a very limited number of frequency samples. The possibility of implementing WTI with a limited number of frequency samples comes from the goal of WTI to reconstruct the low-to-intermediate wavenumbers of the subsurface velocity model using traveltime information. The physical understanding is that seismic waves at different frequencies travel with the same velocity if there is no velocity dispersion during wave propagation.

Time-domain modelling and on-the-fly discrete Fourier transform

The traveltime residual kernel can be computed if the wavefields and Green’s functions are computed and stored beforehand, as indicated in equations (12)–(14).

In our GN-WTI method, the wavefields and Green’s functions are numerically calculated by time-domain finite-difference modelling and are transformed to the frequency domain by on-the-fly discrete Fourier transform (DFT) (Sigure et al., 2008; Xu and McMechan, 2014). For completeness, we now describe the on-the-fly DFT procedure.

In the time domain, we use the finite-difference method to simulate the source-side and receiver-side wavefields. To calculate the frequency-domain wavefields and Green’s functions, we apply on-the-fly DFT at each step of the time-domain extrapolations, as indicated in equation (15):

\[
\tilde{u}(x, \omega; x_i) = \sum_{t=0}^{T} \tilde{u}(x, t; x_i) \exp(i \omega t). \tag{15}
\]

Here, \( \tilde{u}(x, \omega; x_i) \) is the source-side mono-frequency wavefield at an angular frequency \( \omega \) and spatial position \( x \) and \( T \) is the maximum simulation time. Because the forward simulation and on-the-fly DFT are both implemented along the time axis, there is no need to store the entire snapshots at all temporal steps in memory. The receiver-side Green’s functions are calculated in a similar manner.

One valuable advantage of the proposed frequency-domain GN-WTI method is the significant reduction in the computational memory requirement. To provide a detailed comparison of the memory requirements between the straightforward time-domain WTI method by storing entire time-domain wavefields and the proposed frequency-domain GN-WTI method, we considered the following...
simple case: for one shot and one receiver in a three-dimensional (3D) model with $N_x \times N_y \times N_z$ spatial grids, suppose that the step of the time-domain wavefield extrapolation is $N_t$ and the number of selected frequencies for calculating $k_{ij}$ is $N_\omega$. The conventional time-domain WTI method requires $2 \times 4N_x \times N_y \times N_z \times N_t$ memory bytes to store the large floating time-domain wavefield, whereas the proposed frequency-domain GN-WTI method requires $2 \times 8N_x \times N_y \times N_z \times N_\omega$ memory bytes to store the complex source wavefield and Green's function. Therefore, the ratio of the memory requirement between the conventional time-domain WTI method and the proposed frequency-domain GN-WTI method is $r = N_t/(2N_\omega)$. Notably, the proposed frequency-domain GN-WTI method requires less memory than the conventional WTI method as $N_\omega$ is generally significantly smaller than $N_t$. For example, if $N_t = 4000$ and $N_\omega = 10$, then $r = 200$. In particular, the proposed frequency-domain GN-WTI method significantly reduces the memory requirement by more than two orders of magnitude as compared with the conventional time-domain WTI method, which is very promising for the application of WTI in large 3D model inversions.

There are some other techniques for reducing the memory requirement of the time-domain WTI, such as optimal checkpointing methods (Griewank, 1992; Griewank and Walther, 2000; Anderson et al., 2012) and wavefield reconstruction technologies (Dussaud et al., 2008; Clapp, 2009). However, additional wavefield extrapolation steps are needed in these two approaches, resulting in high computational cost.

**COMPARISON OF GAUSS–NEWTON HESSIAN MATRICES FOR RTI AND WTI**

To provide in-depth insight into the WTI Gauss–Newton Hessian matrix (for simplicity, we denote ‘Gauss–Newton Hessian’ as ‘Hessian’ henceforth), we now calculate and compare the Hessian matrices for RTI and WTI in the same acquisition geometry.

First, we calculated the Hessian matrices for a source–receiver pair acquisition geometry. A velocity model with a constant vertical gradient was used to synthesize the observed seismic data. The velocity at the surface was 2000 m/s, and it linearly increased with a gradient of 4 s$^{-1}$. The model was 1.250 km wide in the horizontal direction and 0.625 km in the vertical direction, with a grid spacing of 25 m in both directions ($51 \times 26$ grids in total). The source and the receiver were located at the top surface at $x = 0.050$ km and $x = 1.225$ km, respectively. A Ricker wavelet with a peak frequency of 7 Hz was used in the WTI Hessian matrices calculations. The Hessian matrices for RTI and WTI are shown in Fig. 1(a) and 1(b).

The main diagonal elements of WTI Hessian (Fig. 1b) decayed rapidly with depth (from the upper left corner to the lower right corner) because of the geometrical spreading effect. In contrast, RTI ignores the dynamic characteristics of seismic wave propagation, and the main diagonal elements of its Hessian matrix reflected the length of the rays in the model grids (Fig. 1a). This indicates that WTI Hessian considers the geometrical spreading effect of seismic propagation. The $4 \times 4$ sub-blocks on the upper left corner of the two Hessians in Fig. 1(a) and 1(b) have been enlarged to provide further details, as shown in Fig. 1(c) and 1(d).

These $4 \times 4$ sub-blocks are produced by auto- and cross-correlations of traveltime sensitivity kernel related to the upper-most four layered grids. We select these $4 \times 4$ sub-blocks to demonstrate the finite-frequency band characteristics of the WTI Hessian matrix in detail. The RTI Hessian matrix (Fig. 1c) was composed of sparse discrete points, thereby indicating that the traveltime of one–shot–receiver pair was influenced by the extremely limited spatial grids along the ray path. In contrast, the traveltime of one–shot–receiver pair in WTI was influenced by all spatial grids in the zone around the central ray, which resulted in the finite-frequency band characteristics of the WTI Hessian matrix (Fig. 1d). Therefore, the WTI Hessian can account for the natural defocusing that occurs in response to the limited frequency bandwidth of the seismic data. This indicates that, although both RTI and WTI use seismic wave traveltime information, the wave equation can take into account the finite frequency bandwidth of seismic data and more accurately depict the influence of subsurface velocity on the seismic traveltime. Therefore, WTI can theoretically recover the subsurface velocity model more accurately than RTI.

Afterwards, we calculated the Hessians for a multi-shot and multi-receiver surface seismic acquisition geometry. Here, the velocity model is the same as that in the above example. In this experiment, there were 26 sources regularly distributed on the top surface, and each shot had 26 receivers uniformly distributed on the top surface. The Hessians for RTI and WTI are shown in Fig. 2(a) and 2(b). Figure 2(c) and 2(d) shows the enlarged $4 \times 4$ sub-blocks on the upper left corner of the two Hessian matrices.

As shown by the comparison with the RTI Hessian (Fig. 2a), the main diagonal elements of WTI Hessian (Fig. 2b) decayed more rapidly with depth because of the geometrical spreading effect. For the sub-blocks on the main diagonal (Fig. 2c and 2d), the diagonal dominance with the off-diagonal
bands was clear. Generally speaking, the sub-blocks WTI Hessian (Fig. 2d) exhibited broader off-diagonal bands than those of RTI Hessian (Fig. 2c). On the other hand, the WTI Hessian matrix is smoother than the RTI Hessian matrix. These two phenomena are produced by the finite-frequency band characteristics of the traveltime residual kernel, thus indicating that the WTI Hessian matrix can account for the limited frequency bandwidth of the seismic data. Additionally, these two phenomena indicate that the traveltime residual calculated in WTI can better represent the dependence of the seismic traveltime on the subsurface velocity, although both RTI and WTI use the traveltime information. Therefore, WTI can theoretically recover the subsurface velocity model more accurately than RTI. Moreover, the diagonal elements of WTI Hessian contain
incomplete and uneven illuminations of the seismic acquisition geometry; therefore, application of the inverse Hessian in WTI can be used to filter, or precondition, the gradient vector. Hence, the updating direction can be obtained at a higher accuracy.

**NUMERICAL EXAMPLES**

To demonstrate the superiority of the proposed frequency-domain GN-WTI method, we tested the method using a simple periodic velocity model. The Marmousi model was used to further illustrate its effectiveness. In the inversion process, iterations were conducted until the misfit did not decrease. In both examples, only first arrivals were used and the other waves were muted. To demonstrate the superiority of using Gauss–Newton Hessian information in WTI, the GN-WTI results were compared with the results from WTI using the steepest-descent and L-BFGS approaches. For clarity, we denote WTI using the steepest-descent approach as ‘SD-WTI’ and WTI using the L-BFGS approach as ‘LBFGS-WTI’.

The lateral periodic velocity model

In this section, we apply a lateral periodic velocity model (Fig. 3a) to test the superiority of the frequency-domain GN-WTI method. Figure 3(b) shows the initial model with a constant velocity of 2500 m/s. The model was discretized into 501 × 51 grids with a grid spacing of 10 m in both directions. There were 101 sources evenly distributed on the top surface with a horizontal interval of 50 m, and 101 receivers were regularly fixed on the top surface to record the seismic arrivals. A Ricker wavelet with a peak frequency of 7 Hz was used in the modelling and inversion procedure. The frequencies were selected from 2 to 20 Hz, with a frequency interval of 2 Hz. In this example, the recording length was 2500 ms.

First, the updating directions of the first iteration for different inversion methods were examined. The result showed that the SD-WTI method can attain a correct update direction (steepest-descent direction) (Fig. 4a) but with a very low resolution. Meanwhile, the steepest-descent direction contains obvious acquisition footprints. Compared with the steepest-descent direction (Fig. 4a), the Gauss–Newton direction (Fig. 4b) showed more consistency with the true velocity structure (Fig. 3a). Evidently, GN-WTI can penetrate the deeper part of the model and reach a higher inversion accuracy. Furthermore, the footprints of illumination are eliminated in Fig. 4(b) after the application of the inverse Gauss–Newton Hessian matrix. The preconditioning effect of the Gauss–Newton Hessian in WTI was found to be very significant.

The three WTI methods using different optimization algorithms correctly generated low- to intermediate-wavenumber updates and successfully obtained the near-surface background velocity model (Fig. 5a–5c). Owing to
the application of the inverse Gauss–Newton Hessian, the GN-WTI method yielded better inversion results (Fig. 5c) than the conventional SD-WTI (Fig. 5a) and LBFGS-WTI (Fig. 5b) methods, especially in the deeper part of the model (<0.2 km).

The horizontal velocity profiles at different depths (0, 150, 250 and 350 m below the surface), shown in Fig. 6, demonstrated that the GN-WTI results exhibit higher accuracies at different depths compared with the results of SD-WTI and LBFGS-WTI. At depths of 0 and 150 m, GN-WTI recovered the velocity with higher accuracy. The GN-WTI results fit the true velocity better than the SD-WTI and LBFGS-WTI results. At depths of 250 and 350 m, SD-WTI obtained incorrect updates in many portions, and LBFGS-WTI yielded a better model than SD-WTI (as can be seen from the horizontal velocity profile at a depth of 250 m), but the GN-WTI generated the most accurate velocity updates relatively.

Figure 7 presents the traveltime residuals for different models estimated by cross-correlation functions for the first and 61st shots. The remaining traveltime residuals for the LBFGS-WTI and GN-WTI results were observed to be much smaller than those for the SD-WTI result. The maximum remaining time residual of the SD-WTI result was approximately 11 ms, whereas that of the LBFGS-WTI and GN-WTI results is approximately 2 ms. This indicates that the LBFGS-WTI and GN-WTI results can better explain the first-arrival traveltime than the SD-WTI result. Overall, the remaining traveltime residuals of GN-WTI were smaller than those of LBFGS-WTI, which indicates that GN-WTI produces higher resolutions than LBFGS-WTI.

The final misfit value of the GN-WTI result was much smaller than that of the SD-WTI and LBFGS-WTI results (Fig. 8), which indicates that the GN-WTI result was better at explaining the first-arrival traveltime, as indicated in the previous analysis. Simultaneously, the GN-WTI method also showed the highest convergence rate, as shown in Fig. 8. The misfit of the GN-WTI reached its minimum after only eight iterations. In contrast, the iterations of SD-WTI and LBFGS-WTI were 52 and 25, respectively.

Table 1 shows the comparison between the estimated memory requirement of the conventional time-domain WTI method and that of the frequency-domain GN-WTI method. These data indicate that the GN-WTI method can significantly reduce the computational memory requirement. The proposed method only requires 0.8% of the memory required by the conventional time-domain WTI method. This memory efficiency advantage of the GN-WTI method will be of great importance in real, large 3D applications.

On the other hand, the computational cost of the GN-WTI method was found to be in the same order of magnitude as that of SD-WTI and LBFGS-WTI methods in this example, since no additional forward modelling is needed to compute Hessian vector products. The total CPU time of the GN-WTI method was 8320 s, whereas the total CPU times of SD-WTI and LBFGS-WTI were 8448 and 4250 s, respectively. Although the CPU time for one iteration of the GN-WTI was approximately 6.5 times that of SD-WTI, the total CPU time of the GN-WTI was slightly lower than that of SD-WTI, which
An efficient Gauss–Newton WTI method

Figure 6 Horizontal velocity profiles of inverted velocity models at depths of 0 m (a), 150 m (b), 250 m (c) and 350 m (d). The solid black and dashed black lines indicate the true and initial models, respectively. The dotted black, solid blue and solid red lines indicate the SD-WTI, LBFGS-WTI and GN-WTI results, respectively.

Figure 7 Remaining first-arrival traveltime residuals estimated by cross-correlation functions for SD-WTI (black), LBFGS-WTI (blue) and GN-WTI (red) results of the first (a) and 61st (b) shots.

Figure 8 Convergence curves of misfit functions for SD-WTI (black), LBFGS-WTI (blue) and GN-WTI (red) methods.

was a result of the high convergence rate obtained in response to the application of the inverse Gauss–Newton Hessian.

The Marmousi model

Next, we examine the advantages of the frequency-domain GN-WTI method using the more complex Marmousi model (Fig. 9a). In this example, WTI and FWI were combined to retrieve high-resolution results. As the initial velocity model (Fig. 9b) was not accurate (i.e., it increased linearly with depth), different WTI methods were used to obtain initial models for subsequent FWI. The frequency band used in the WTI was [2 Hz, 20 Hz], with a frequency sampling interval of 2 Hz. In both WTI and FWI implementations, the recording length was 5000 ms. The model was 7.5 km wide in the horizontal direction and 2.1 km in the vertical direction, with a grid spacing of 15 m in both directions. A total of 101 shots...
were regularly deployed along the top surface of the model. Each shot was recorded by a receiver line with a spacing of 75 m, and the maximum offset was 6 km. A Ricker wavelet with a peak frequency of 7 Hz was used.

Figure 10(a) and 10(b) shows the steepest-descent direction and Gauss–Newton direction for WTI of the first iteration, respectively. The SD-WTI method obtained the correct updating direction (steepest-descent direction) (Fig. 10a), whereas the resolution was rather low and the deep part below 1.2 km was not illuminated. An apparent velocity anomaly exists near the top surface. This true velocity anomaly was not observed in the steepest-descent direction, but it was evident in the Gauss–Newton direction (as indicated by the red circle in Fig. 10b). Meanwhile, the steepest-descent direction contained footprints of the seismic acquisition geometry, which may lead to undesired updates during the inversion. Compared with the steepest-descent direction, the Gauss–Newton direction of the first iteration was found to be more consistent with the true velocity structure, and it correctly updated the deep part of the model and reached a higher resolution (Fig. 11c). Evidently, the undesired footprints of illumination were eliminated.
after the application of the inverse Gauss–Newton Hessian matrix. The filtering effect of the Gauss–Newton Hessian was significant.

The final SD-WTI, LBFGS-WTI and GN-WTI results after 20, 21 and 9 iterations are shown in Fig. 11. The SD-WTI mainly updated the shallow part of the velocity model with a maximum depth of approximately 1.4 km. The LBFGS-WTI correctly updated the deeper part of the model, but the improvement was limited. On the other hand, the GN-WTI correctly modified the velocity model in the deeper part to a maximum depth of approximately 2 km. Apparently, the GN-WTI method can yield better inversion results than the conventional SD-WTI and LBFGS-WTI methods, especially in the deep part. Vertical velocity profiles at $x = 3.0$ and $5.25$ km (Fig. 12) indicated that the SD-WTI, LBFGS-WTI and GN-WTI methods reconstruct nearly the same background velocity from the top surface to 1 km. However, in deeper parts (1–2 km), GN-WTI exhibit significant accuracy improvements and reconstructed a more accurate velocity model.

To evaluate the quality of the inverted results, the root-mean-square velocity model error $e_{\text{rms}}$ was calculated according to equation (16):

$$e_{\text{rms}} = \sqrt{\frac{1}{n} \sum \left( \frac{v_{\text{inv}} - v_{\text{true}}}{v_{\text{true}}} \right)^2},$$

where $v_{\text{true}}$ and $v_{\text{inv}}$ denote the true and inverted velocities, respectively. Figure 13(a) shows the root-mean-square model errors at different iterations for the SD-WTI, LBFGS-WTI and GN-WTI methods. The final inverted error of the GN-WTI result was much smaller than that of the LBFGS-WTI and SD-WTI results. This indicated that the velocity was better reconstructed by using the GN-WTI method. The misfit convergence curves in Fig. 13(b) show that the SD-WTI method converged after 20 iterations and the LBFGS-WTI method converged to a smaller value after 21 iterations, whereas the GN-WTI method converged to a much smaller misfit value.
after only nine iterations. This proves that the GN-WTI leads to a higher convergence rate and can better interpret the first-arrival traveltime.

The WTI methods can provide initial velocity models with low-wavenumber velocity information for subsequent FWI to further improve the model accuracy. Figure 14(a) shows the FWI inverted velocity model with the SD-WTI final result as the initial velocity model. Because the seismic data contained some long-offset refracted waves, and the SD-WTI result was similar to the exact model in the shallow part, the conventional FWI produced an acceptable inversion result at the shallow part. However, the presence of errors in the deep part of the SD-WTI result led to incorrect updates in the deeper part of the FWI result. The inverted model (Fig. 14b) by FWI with LBFGS-WTI result as the initial model was found to be better than that with the SD-WTI result, but this model also became stuck in a local minimum. As the GN-WTI method provided a better initial velocity model, the final FWI result (Fig. 14c) revealed most of the major features, and thus, these results were superior to those shown in Fig. 14(a) and 14(b).

The comparison of memory requirements between the frequency-domain GN-WTI method and conventional time-domain WTI method in this example was similar to the above example because the memory reduction ratio mainly depends on the ratio of \( N_t \) and \( N_\omega \). We focused on the analysis of the computational cost in this WTI example. The total CPU time of the GN-WTI method was 5270 s, whereas the total CPU times of SD-WTI and LBFGS-WTI were 6479 and 4343 s, respectively. The total CPU time of the GN-WTI was lower than that of SD-WTI owing to the high convergence rate after application of the inverse Gauss–Newton Hessian. The GN-WTI method had a slightly longer CPU time than that of the LBFGS-WTI method, but it obtained more accurate results. This WTI example further validates the finding that the computational cost of the proposed GN-WTI method is on the same order of magnitude as that of the SD-WTI and LBFGS-WTI methods.

**DISCUSSION**

In this study, the gradient and Hessian vector products were computed by the improved SI method. The improved SI method is preferred when the number of sources is equal to or more than a few percent of the number of non-redundant receivers, a condition that is common in controlled-source exploration, ocean bottom seismic exploration and earthquake seismology. Some special seismic surveys may also meet this condition, such as the two WTI examples in this paper. The number of non-redundant receivers was zero in the two WTI examples since receiver locations coincided with the source locations, which made the improved SI approach much more efficient than the adjoint-state approach in our proposed GN-WTI method. The dense receiver spacing in seismic surveys may make the SI-based GN-WTI method computationally expensive.

**CONCLUSION**

We propose a truncated GN-WTI method implemented in the frequency domain to retrieve the background near-surface velocity model. The comparison of Hessian matrices for RTI and WTI indicated that WTI can recover the subsurface velocity model more accurately than RTI, although both of them...
use the traveltime information. The memory requirement of the proposed frequency-domain GN-WTI method is greatly reduced compared with the conventional time-domain WTI method because only a few frequencies are used for inversion. The Hessian vector products are efficiently computed using the improved SI method as long as the source-side wavefields and non-redundant receiver-side Green’s functions are computed and stored in advance. Two numerical WTI examples validated the conclusion that the computational cost of the proposed GN-WTI method is on the same order of magnitude as that of SD-WTI and LBFGS-WTI methods owing to the high convergence rate after application of the inverse Gauss–Newton Hessian. The numerical example of a lateral periodic velocity model also demonstrated the superiority of the proposed GN-WTI method in terms of the convergence rate and inversion accuracy compared with the SD-WTI and LBFGS-WTI methods. The application to the complex Marmousi model further illustrated the effectiveness of the proposed method in complicated situations. In summary, the proposed method should be beneficial in near-surface velocity model building.

ACKNOWLEDGEMENTS

We thank anonymous reviewers and editors for their detailed and constructive comments. We are thankful for the financial support from the National Key R&D Program of China (grant number: 2019YFC0312004) and the National Natural Science Foundation of China (grant number: 41874127).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data associated with this research are available and can be obtained by contacting the corresponding author.

ORCID

Jianhua Wang https://orcid.org/0000-0002-7554-7794

REFERENCES

Anderson, J.E., Tan, L. and Wang, D. (2012) Time-reversal checkpointing methods for RTM and FWI. Geophysics, 77(4), S93–S103.

Cova, R., Henley, D. and Innanen, K.A. (2018) Computing near-surface velocity models for S-wave static corrections using raypath interferometry. Geophysics, 83(3), U23–U34.

Clapp, R.G. (2009) Reverse time migration with random boundaries. 79th Annual International Meeting, SEG, Expanded Abstracts, 2809–2813.

Dussaud, E., Symes, W.W., Williamson, P., Lemaistre, L., Singer, P., Denel, B. and Cherrett, A. (2008) Computational strategies for reverse-time migration. 78th Annual International Meeting, SEG, Expanded Abstracts, 2267–2271.

Griewank, A., (1992) Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation. Optimization Methods and Software, 1, 35–54.

Griewank, A. and Walther, A. (2000) Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation. ACM Transactions on Mathematical Software, 26, 19–45.

Keho, T.H. and Zhu, W. (2009) Revisiting automatic first arrival picking for large 3D land surveys. 79th Annual International Meeting, SEG, Expanded Abstracts, 3198–3202.

Liu, Y., Yang, J., Chi, B. and Dong, L. (2015) An improved scattering-integral approach for frequency-domain full waveform inversion. Geophysical Journal International, 202, 1827–1842.

Liu, Y., Wu, Z., Kang, H. and Yang, J. (2020) Use of prismatic waves in full-waveform inversion with the exact Hessian. Geophysics, 85(4), R325-R337.

Luo, Y. and Schuster, G.T. (1991) Wave-equation traveltime inversion. Geophysics, 56, 645–653.

Luo, Y., Ma, Y., Wu, Y., Liu, H. and Cao, L. (2016) Full traveltime inversion. Geophysics, 81(5), R261-R274.

Ma, Y. and Hale, D. (2013) Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion. Geophysics, 78, R223–R233.

Métivier, L., Brossier, R., Virieux, J. and Operto, S. (2013) Full waveform inversion and the truncated Newton method. SIAM Journal on Scientific Computing, 35(2), B401-437.

Métivier, L., Bretaudeau, F., Brossier, R., Operto, S. and Virieux, J. (2014) Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures. Geophysical Prospecting, 62(6), 1353–1375.

Nocedal, J. and Wright, S.J. (2006) Numerical Optimization, 2nd edition. Springer Series in Operations Research and Financial Engineering. Springer Science + Business Media, LLC.

Olson, K.B. (1989) A stable and flexible procedure for the inverse modeling of seismic first arrivals. Geophysical Prospecting, 37, 455–465.

Sei, A. and Symes, W.W. (1994) Gradient calculation of the traveltime cost function without ray tracing. 64th Annual International Meeting, SEG, Expanded Abstracts, 1351–1354.

Sheriff, R.E. (2002) Encyclopedia Dictionary of Applied Geophysics. Society of Exploration Geophysicists

© 2021 The Authors. Near Surface Geophysics published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists and Engineers., Near Surface Geophysics, 19, 415–428.
Tape, C., Liu, Q., Maggi, A. and Tromp, J. (2010) Seismic tomogra-
phy of the southern California crust based on spectral-element and
adjoint methods. *Geophysical Journal International*, 180, 433–462.
Van leeuwen, T. and Mulder, W.A. (2010) A correlation-based mis-
fit criterion for wave-equation traveltime tomography. *Geophysical
Journal International*, 182, 1383–1394.
Vigh, D., Starr, E.W. and Kapoor, J. (2009) Developing earth model
with full waveform inversion. *The Leading Edge*, 28, 432–435.
Virieux, J. and Operto, S. (2009) An overview of full-waveform inver-
sion in exploration geophysics. *Geophysics*, 74(6), WCC1–WCC26.
Wang, B., Kim, Y.S., Mason, C. and Zeng, X. (2008) Advances in ve-
locity model-building technology for subsalt imaging. *Geophysics*,
73, no. (5), VE173-VE181.
Woodward, M.J. (1992) Wave-equation tomography. *Geophysics*,
57(1), 15–26.
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.
Yang, J., Liu, Y. and Dong, L. (2016) Simultaneous estimation of veloc-
ity and density in acoustic multiparameter full-waveform inversion
using an improved scattering-integral approach. *Geophysics*, 81(6),
R399–415.
Zelt, C.A., Azaria, A. and Levander, A. (2006) 3D seismic refraction
traveltime tomography at a groundwater contamination site. *Ge-
ophysics*, 71(5), H67-H78.
Zhang, J. and Toksöz, M.N. (1998) Nonlinear refraction traveltime
tomography. *Geophysics*, 63, 1726–1737.
Zhao, D., Yanada, T., Hasegawa, A., Umino, N. and Wei, W. (2012)
Imaging the subducting slabs and mantle upwelling under the Japan
Islands. *Geophysical Journal International*, 190, 816–828.
Zheng, Y., Wang, Y. and Chang, X. (2013) Wave-equation traveltime
inversion: comparison of three numerical optimization methods.
*Computers & Geosciences*, 60, 88–97.
Zhou, C., Cai, W., Luo, Y., Schuster, G.T. and Hassanzadeh, S. (1995)
Acoustic wave-equation traveltime and waveform inversion of
crosshole seismic data. *Geophysics*, 60(3), 765–773.
Zhu, X., Sixta, D.P. and Angstman, B.G. (1992) Tomostatics: turning-
ray tomography + static corrections. *The Leading Edge*, 11, 15–23.

APPENDIX

**WORKFLOW AND IMPLEMENTATION OF THE GN-WTI METHOD**

The workflow and implementation of our GN-WTI method in the
frequency domain are presented below.
1. Construct the initial velocity model \( v_0 \) and input the ob-

served common-shot gathers.
2. Calculate the frequency-domain source wavefields
\( u(x, \omega; x_s) \) and Green’s functions \( G(x, \omega; x_r) \) of all sources

and all non-redundant receivers by time-domain forward
modelling and on-the-fly DFT (Sirgue et al., 2008). Af-

ter the time-domain forward modelling for all sources

is finished, the synthetic common-shot gathers are also

obtained.
3. Extract the observed first arrivals \( d_{\text{obs}}(x_s, x_r; t) \) and the
calculated first arrivals \( d_{\text{cal}}(x_s, x_r; t) \) by applying window
functions to the common-shot gathers. Construct the travel-
time residual vector \( \Delta t \) and calculate the value of the misfit
function \( J \) in equation (1). Check the misfit function and the
stopping criteria. Either stop or go to the next iteration.
4. Calculate the traveltime residual kernel according to equa-
tions (12)–(14).
5. Calculate the gradient of the misfit function \( \nabla v J \) accord-

ing to equation (8).
6. Calculate the Gauss–Newton direction \( p_{\text{GN}} \) by solving

equation (4) using a matrix-free CG algorithm (Nocedal and

Wright, 2006).
7. Search a suitable step-length by parabolic interpolation

(Vigh et al., 2009).
8. Update the model \( v_{k+1} = v_k + \gamma_k p_{\text{GN}}^k \).
9. Repeat steps 2–8.