ORIGINAL PAPER

3D variable-grid full-waveform inversion on GPU

Zi-Ying Wang[1] · Jian-Ping Huang[1] · Ding-Jin Liu[2] · Zhen-Chun Li[1] · Peng Yong[1] · Zhen-Jie Yang[1]

Abstract

Full-waveform inversion (FWI) is a powerful tool to reconstruct subsurface geophysical parameters with high resolution. As 3D surveys become widely implemented, corresponding 3D processing techniques are required to solve complex geological cases, while a large amount of computation is the most challenging problem. We propose an adaptive variable-grid 3D FWI on graphics processing unit devices to improve computational efficiency without losing accuracy. The irregular-grid discretization strategy is based on a dispersion relation, and the grid size adapts to depth, velocity, and frequency automatically. According to the transformed grid coordinates, we derive a modified acoustic wave equation and apply it to full wavefield simulation. The 3D variable-grid modeling is conducted on several 3D models to validate its feasibility, accuracy and efficiency. Then we apply the proposed modeling method to full-waveform inversion for source and residual wavefield propagation. It is demonstrated that the adaptive variable-grid FWI is capable of decreasing computing time and memory requirements. From the inversion results of the 3D SEG/EAGE overthrust model, our method retains inversion accuracy when recovering both thrust and channels.

Keywords 3D full-waveform inversion · Adaptive variable grid · Finite-difference modeling

1 Introduction

A high-resolution velocity model is key to migration methods and reservoir estimation. Traveltime tomography has been widely used in velocity model reconstruction, but it is based on the high-frequency assumption and has low resolution. Thanks to the considerable progress of computation ability, FWI has become a potential method to reveal detailed information by iteratively updating subsurface model parameters, the misfit between observed data and synthetic data minimized using a gradient-based method (Virieux and Operto 2009). Tarantola (1984) proposed the adjoint state method to estimate the gradient of misfit by a zero-lag correlation between the source wavefield and the residual wavefield, which avoids explicitly calculating the Jacobi matrix. Bunks et al. (1995) developed a multi-scale inversion strategy. Over the past decade, a lot of research has been conducted to solve FWI problems, such as the "cycle skipping" problem (Shin and Cha 2008; Van Leeuwen and Mulder 2010; Ma and Hale 2013; Métivier et al. 2016) and expensive computational costs (Ben-Hadj-Ali et al. 2011).

The needs of 3D seismic techniques have been increasing recently as more 3D wide-azimuth surveys have been carried out; however, the processing demands prohibitive computation costs. Ben-Hadj-Ali et al. (2008) and Sirgue et al. (2008) developed 3D FWI and provided preliminary insights on the feasibility and advance of 3D FWI in the time domain and frequency domain, respectively, based on some synthetic tests. Further 3D tests on field data sets were presented (Plessix and Perkins 2010; Warner et al. 2013; Raknes et al. 2015). In order to reduce computation, the simultaneous-shot technique provides a trade-off between computational efficiency and imaging quality, introducing cross-talk noises (Ben-Hadj-Ali et al. 2011). Vigh and Starr (2008) studied 3D prestack plane-wave FWI that reduced the amount of computation. Since the Compute Unified Device Architecture (CUDA) programming language was employed in computational science, graphics processing units (GPUs) have been

✉ Jian-Ping Huang
jphuang@upc.edu.cn

[1] School of Geosciences, China University of Petroleum, Qingdao 266580, Shandong, China

[2] Sinopec Geophysical Research Institute, Nanjing 211103, Jiangsu, China

widely applied to 3D data in exploration geophysics. It is presented that GPU runtimes are roughly one-tenth to one-twentieth of corresponding multi-core CPU-based implementations (Weiss and Shragge 2013). For example, 3D Laplace-domain full-waveform inversion on a single GPU card was verified to be accurate when used to invert a long-wavelength velocity model, and the peak speedup of GPU parallelization is 24.6 times (Shin et al. 2014). Liu et al. (2015) achieved a good acceleration ratio up to 19.5 using 3D simplified hybrid-domain FWI on GPU. The time-domain FWI is less memory-demanding but time-consuming; hence, intensive computation time still remains a distinct weakness of 3D time-domain FWI (Jiang and Zhu 2018). When tens of iterations and hundreds of shots are performed to characterize subsurface velocity structures, an efficient wavefield numerical modeling algorithm is the core of accelerating time-domain FWI (Vigh and Starr 2008).

During the process of FWI, forward propagation of the source wavefield and backward propagation of the residual wavefield occupy the largest portion of operation time. Therefore, the numerical simulation algorithm is the key to promoting efficiency. The variable-grid technique changes grid size in different areas aiming at near-surface low-velocity zone and complicated small-scale anomalies in deeper layers. The advantage of this method is that the improvement in efficiency, the reduction of memory requirements and the promotion of precision can be achieved simultaneously. Irregular spatial grid size was firstly proposed by Moczo (1989). Jastram and Behle (1992) modeled on a grid of varying spacing in depth based on a two-dimensional acoustic wave equation. Jastram and Tessmer (1994) developed it into elastic cases. Zhu et al. (2007) and Sun et al. (2008) proved that the variable-grid method in depth promotes the accuracy and efficiency of finite-difference forward modeling, and error was evaluated and reduced by changing the difference operator. Huang et al. (2015) combined the coordinate transformation for the irregular surface and a mapping staggered-grid finite-difference forward modeling method using the dual-variable-grid size in time and space. Zhang et al. (2018) used a discrete fracture model (DFM) to reflect the shape and distribution of the fractures in the flow model and to simulate reservoir production with different fracture parameters. Fan et al. (2015) applied a rotated grid system to the transition region for high-order finite-difference modeling that is capable of doubling the grid spacing without artificial reflections. Fan et al. (2018) further developed a discontinuous-grid finite-difference scheme for frequency-domain 2D scalar wave modeling. However, the fine-to-coarse spacing ratio of both modeling methods (Fan et al. 2015, 2018) is restricted to a power of two. To double the grid size, a transition zone in the coarse-grid area is indispensable. The

modeling algorithms and strategies mentioned above showed good performance on models with a shallow low-velocity layer and a complex subsurface geological body, such as gas clouds and small-scale low-velocity inclusions.

Variable-grid modeling techniques have been used in migration and waveform inversion. Ha and Shin (2012) used an axis transformation technique to improve the Laplace-domain FWI efficiency, in which two transformation functions are proposed. Li et al. (2014) applied a dual-variable-grid forward modeling algorithm with high precision and efficiency to reverse time migration, which improves the accuracy in fractured reservoirs and irregular surface imaging. The dual-variable-grid method in time and space was applied to encoding FWI, and a multi-scale strategy was realized by reducing the global grid scale and fine grid scale (Qu et al. 2015). Frequency-domain FWI based on a variable-grid finite-difference method was also implemented to prove its feasibility and effectiveness by numerical tests (Li et al. 2016). However, a conventional grid discretization strategy always finely resamples a specified zone whose edge creates artificial reflections because the grid size in this zone is several times smaller than the global ones. Such an artificial error is difficult to remove from the calculated wavefield. In addition, all the above attempts are limited to 2D cases while the enormous computation cost, the biggest problem in 3D FWI, urgently needs to be reduced. Up to the present, variable-grid methods have not been used in 3D FWI on GPU devices to test their feasibility, efficiency and accuracy.

In this paper, we propose a 3D optimized variable-grid FWI on a single GPU card and 3D numerical tests are operated to verify the validity and efficiency of the method. Based on the dispersion relation, grid sizes adapt to local velocity and depth, which not only simulates wavefields with higher accuracy but also accelerates the calculation during the implementation of FWI. With the obvious reduction of computing time, the proposed FWI method inverts for the velocity model successfully. In the following sections, we will firstly review the FWI theory. Then we will introduce our grid discretization strategy, deriving a new wave equation used in adaptive variable-grid FWI. The GPU implementation workflow of this method is also detailed in the next part. After that, 3D numerical test results are shown using simple models to demonstrate the efficiency and accuracy of the proposed modeling method. In the next part, a complex 3D model, the 3D SEG/EAGE overthrust model, is used to show that application of this method to FWI is feasible. From the inverted results and computing time comparison, variable-grid FWI shows higher computational efficiency and has better convergence. In the last section, conclusions will be presented.

2 Theory

2.1 Adaptive variable-grid modeling

Time-domain 3D FWI is challenging due to the large amount of calculation. Conventional FWI usually uses a fixed regular grid while modeling the forward wavefield and the backward propagating residual wavefield. In most cases, the fixed global grid size is not fine enough in the near-surface low-velocity zone. Besides, in deeper layers with higher velocity, the model is oversampled and the calculation cost is increased without increasing accuracy. To improve this situation, we applied adaptive variable-grid modeling to FWI, which achieves higher efficiency without losing accuracy. In this section, we will focus on presenting the theory and strategy of adaptive variable-grid modeling.

The 3D time-domain acoustic wave equation with constant density with regular-grid discretization can be written as

$$\frac{1}{v^2(\mathbf{x})}\frac{\partial^2 u}{\partial t^2} = \left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right] + f_s(t) \tag{1}$$

where $v(\mathbf{x})$ is the velocity, $f_s(t)$ is the source, and u is the pressure field; x, y, z are the position in the regular-grid coordinate.

As shown in Fig. 1a, the spatial sampling intervals dx, dy, dz are fixed in the direction of x, y, z. When the near-surface velocity is low or low-velocity targets exist in the deeper zone, such coarse-grid discretization will lead to inaccurate forward and backward wavefields due to dispersion phenomenon. On the other hand, if we use fixed fine

grids, the large computing and memory storage problem will be more severe especially for the implementation of FWI. Moreover, conventional variable-grid methods are fixed local grid refinement methods, which can only refine a specified area. The abrupt change of grid size in the transitional zone from fine grid to coarse-grid results in artificial reflections.

Here, we use a new grid discretization strategy to resample velocity models in the z direction, based on dispersion relation written as

$$\lambda_{\min} = \frac{v_{\min}}{f_{\max}} \tag{2}$$

where $\lambda_{\min}$ denotes the minimum wavelength, $v_{\min}$ represents the minimum velocity in the model, and $f_{\max}$ represents the maximum frequency of the source signal. Usually, the maximum frequency of the Ricker wavelet is three times the wavelet dominant frequency. So we refine the z-axis using the following equation:

$$dz = \frac{v(z)}{nf_d} \tag{3}$$

where $v(z)$ denotes the velocity in depth z. To ensure dispersion never occurs, we use the minimum velocity among all grid points in that depth to replace $v(z)$. f_d represents the dominant frequency of the Ricker wavelet. n is the number of grid points per wavelength, which is set as ten in this work to ensure the wavefields are adequately resampled. The horizontal grid size dx and dy is fixed and equals to the regular-grid size, while the vertical grid size dz adapts to the local velocity and automatically changes with depth. This is
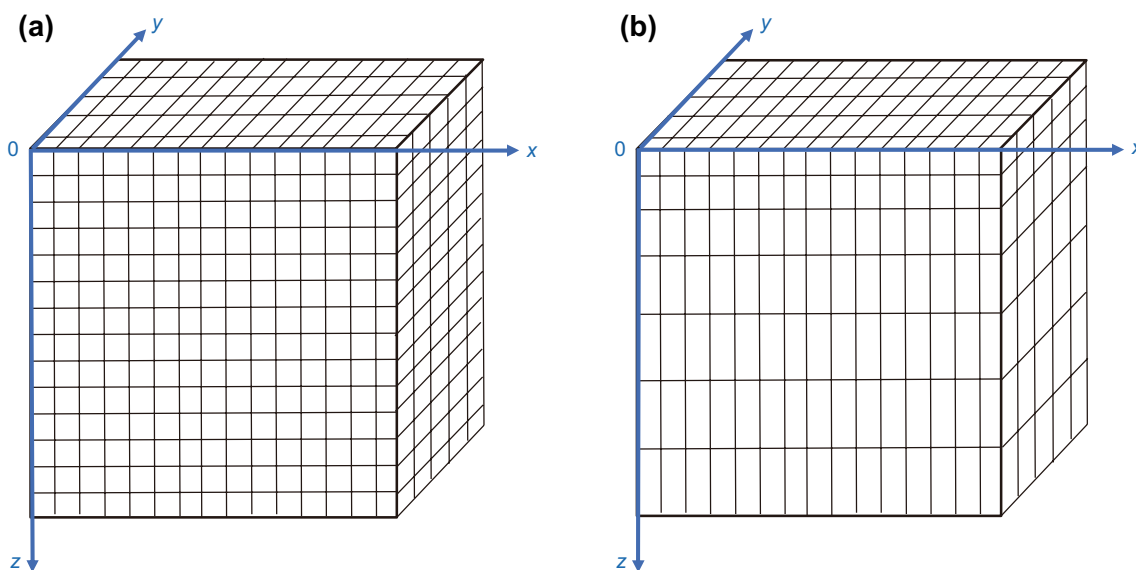

Fig. 1 Physical coordinate with traditional (**a**) and adaptive variable-grid discretization strategy (**b**)

because the main velocity changing direction is along the depth direction, especially for background velocity models.

Using the dispersion relation mentioned above, we can calculate theoretical dz on each grid point according to its local velocity. With the theoretical grid intervals obtained, we take trial steps from $z = 0$ to get the first adaptive-grid depth like the five blue rectangles in Fig. 2. When the trial step length is equal to z (the first red square in Fig. 2), we will get the first feasible grid point z_1. Then we restart the attempt with a new trial step until we get the next square and z_2. We repeat the process to the maximum depth, and we obtain the depth of each optimal grid point z_i; in other words, we get the new grid discretization with the adaptive variable grid. Note that the process of taking trial steps is necessary, because if we simply use theoretical dz as grid size and the local velocity is decreasing, it is not fine enough to assure that modeling is dispersion free. After that, we smooth z_i to prevent spurious reflections. In this way, the regular physical coordinate is transformed into a new physical coordinate with variable vertical grid intervals.

The resampled physical coordinate is shown in Fig. 1b. From this grid discretization graph, we can see that the lateral grid size is never changed. The vertical grid size increases gradually from shallow layers with low velocity to deep layers with high velocity. In general cases, the vertical grid size at the bottom of velocity models amounts up to several times the fine grid size. The new discretization strategy enables the grid size to increase gradually according to depth and velocity, instead of a violent change in the transitional zone that usually amounts to several times the fine grid size. That is the reason why our method has higher accuracy than the conventional variable-grid method.

Therefore, we have two coordinate systems and their transformation relation can be written as:

$$x_{vg} = x$$
$$y_{vg} = y \qquad (4)$$
$$z_{vg} = \varphi(z)$$

where x_{vg}, y_{vg}, and z_{vg} are the new variable-grid location in the transformed coordinate system; $\varphi(z)$ is the mapping from the original depth to the new depth after spatial resampling. Compared with a fixed axis transformation function proposed by Ha and Shin (2012), this function $\varphi(z)$ is capable of varying according to different velocity models. The new axis transformation function $\varphi(z)$ is changeable under different geological circumstances. Moreover, errors originating from large grids in deep layers are also avoided. In this method, the coordinate only changes in depth while the lateral grid size is fixed. Therefore, the x, y position of grid points is identical to the original ones. Here, we will focus on the formula derivation in the direction of z.

We can get the first-order and second-order derivatives of z_{vg} with respect to the original physical coordinate variables z using a finite-difference method:

$$\frac{\partial z_{vg}}{\partial z} = \varphi'(z) \qquad (5)$$

According to the derivation of inverse functions, we derive the derivatives of z, with respect to z_{vg}:

$$\frac{\partial z}{\partial z_{vg}} = \frac{1}{\varphi'(z)} \qquad (6)$$

$$\frac{\partial z^2}{\partial z_{vg}^2} = \frac{\partial \left(\frac{1}{\varphi'(z)} \right)}{\partial z} \frac{\partial z}{\partial z_{vg}} = \frac{-\varphi''(z)}{[\varphi'(z)]^2} \frac{1}{\varphi'(z)} = -\frac{\varphi''(z)}{[\varphi'(z)]^3} \qquad (7)$$

Based on Eq. 1, we derive the acoustic wave equation in the new coordinate system. The first-order derivative of the replacement is written as:

$$\frac{\partial u}{\partial z_{vg}} = \frac{\partial u}{\partial x} \frac{\partial x}{\partial z_{vg}} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial z_{vg}} + \frac{\partial u}{\partial z} \frac{\partial z}{\partial z_{vg}} = \frac{\partial u}{\partial z} \frac{\partial z}{\partial z_{vg}} \qquad (8)$$

Also, we can derive the second derivative:

$$\begin{aligned}
\frac{\partial^2 u}{\partial z_{vg}^2} &= \frac{\partial \frac{\partial u}{\partial z_{vg}}}{\partial x} \frac{\partial x}{\partial z_{vg}} + \frac{\partial \frac{\partial u}{\partial z_{vg}}}{\partial y} \frac{\partial y}{\partial z_{vg}} + \frac{\partial \frac{\partial u}{\partial z_{vg}}}{\partial z} \frac{\partial z}{\partial z_{vg}} \\
&= \left(\frac{\partial^2 u}{\partial z^2} \frac{\partial z}{\partial z_{vg}} + \frac{\partial \frac{\partial z}{\partial z_{vg}}}{\partial z} \frac{\partial u}{\partial z} \right) \frac{\partial z}{\partial z_{vg}} \\
&= \frac{\partial^2 u}{\partial z^2} \left(\frac{\partial z}{\partial z_{vg}} \right)^2 + \frac{\partial u}{\partial z} \frac{\partial^2 z}{\partial z_{vg}^2} \\
&= \frac{\partial^2 u}{\partial z^2} \left(\frac{1}{\varphi'(z)} \right)^2 - \frac{\partial u}{\partial z} \frac{\varphi''(z)}{[\varphi'(z)]^3}
\end{aligned} \qquad (9)$$

Fig. 2 Variable-grid discretization strategy to obtain transformed z_{vg}

Table 1 Computing time and speedup of modeling on GPU

Number of grids	$100 \times 100 \times 100$	$200 \times 200 \times 200$	$300 \times 300 \times 300$
Computing time on CPU (1 core), s	127.9	1272.2	4363.0
Computing time on GPU, s	1.3	7.7	13.1
Speedup	98.4	165.2	333.1

To sum up, we can rewrite Eq. 1 as follows:

$$
\frac{1}{v^2(\mathbf{x})} \frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x_{\mathrm{vg}}^2} + \frac{\partial^2 u}{\partial y_{\mathrm{vg}}^2} + \frac{\partial^2 u}{\partial z_{\mathrm{vg}}^2}
$$
$$
= \left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \left(\frac{1}{\varphi'(z)} \right)^2 - \frac{\partial u}{\partial z} \frac{\varphi''(z)}{[\varphi'(z)]^3} \right] + f_s(t)
$$
(10)

Equation 10 is the modified wave equation for variable-grid modeling. One term added to Eq. 1 increases the computation complexity for one grid point. Nevertheless, this small amount of extra calculation can be neglected compared with the large reduction of the calculated number of grid points. For 3D time-domain acoustic modeling, we adopt an 8th-order finite-difference operator. For the boundary condition, we applied an absorbing boundary condition suggested by Clayton and Engquist (1977) to minimize artificial reflected waves from the edges of the computation domain.

2.2 Full-waveform inversion based on GPU acceleration

As the most intensive and time-consuming step, full wavefield simulation is accelerated using the proposed grid discretization strategy. A GPU device has higher performance than CPU devices in accelerating the calculation for the data-parallel tasks in that it has several hundred or more thread processors (Jiang and Zhu 2018; Yang et al. 2015). The large computation cost is a challenging problem of 3D full-waveform simulation with the performance of a number of shots. Therefore, it is beneficial to apply GPU parallelization techniques in the procedural steps. To further increase efficiency, we use the CUDA-C language for GPU programming on a workstation with a single GPU card. The GPU card used in this paper is the NVIDIA Quadro P5000, which has 2560 processors, and 16 GB of GDDR5X memory. It uses PCI-E to connect with an Intel Xeon E5-2630 v4 CPU running at 2.20-GHz with 256 GB RAM memory. To illustrate the performance increase in GPU implementation, we test the efficiency of the CUDA-C code and CPU code based on several 3D models with different grid numbers (Table 1).

We parallelize key computation modules of 3D FWI, mainly including forward modeling, backward propagating residual wavefield, estimating descent direction and searching step length. Each step above is calculated in a GPU kernel function. Resampling the velocity model is not as computationally intensive as the other steps, and hence it is implemented on the CPU, including obtaining $\varphi(z)$ and calculating corresponding derivatives. Moreover, linear interpolation is applied to transform the wavefield snapshots and inversion results to the original regular-grid coordinate. The inverse transformation of the inverted velocity model works at the final step rather than after each full wavefield simulation in each shot loop. The working flowchart of the variable-grid FWI is shown in Fig. 3, in which the calculation modules with a gray background are implemented on the GPU.

FWI estimates subsurface model parameters by minimizing the least-squares distance between the synthetic and observed data. The L_2 norm difference objective function is defined as (Tarantola 1984),

$$
E(\mathbf{m}) = \frac{1}{2} \sum_{\mathrm{s,r}} \int_0^{t_{\max}} [u(\mathbf{x}_{\mathrm{r}}, t; \mathbf{x}_{\mathrm{s}}) - u_0(\mathbf{x}_{\mathrm{r}}, t; \mathbf{x}_{\mathrm{s}})]^2 \mathrm{d}t
$$
(11)

where $u(\mathbf{x}_{\mathrm{r}}, t; \mathbf{x}_{\mathrm{s}})$ is synthetic data; $u_0(\mathbf{x}_{\mathrm{r}}, t; \mathbf{x}_{\mathrm{s}})$ is observed data recorded at time t; $\mathbf{x}_{\mathrm{s}}$ and $\mathbf{x}_{\mathrm{r}}$ are the locations of sources and receivers; and $t_{\max}$ is the largest calculated time.

FWI is a large-scale nonlinear optimization problem, and we generally use a gradient-based local optimization method to minimize the objective function and update the model parameters iteratively. Starting with an initial model $\mathbf{m}_0$, the model is updated iteratively as follows:

$$
\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{d}_k
$$
(12)

where $\mathbf{m}_k$ is the updated model at the kth iteration; α_k is the step length obtained by parabolic interpolation to minimize the objective function in the descent direction $\mathbf{d}_k$.

In practice, the gradient of the objective function $E(\mathbf{m})$ with respect to the model parameter is derived through the adjoint state method (Plessix 2006; Fichtner and Trampert 2011). It is formulated as the correlation of the source wavefield forward propagated from the source with the adjoint wavefield back propagated from the residual at the receivers (Tarantola 1984), which is expressed as:

$$
\nabla E_{\mathbf{m}} = \frac{2}{v^3(\mathbf{x})} \sum_{\mathrm{s,r}} \int_0^{t_{\max}} \frac{\partial^2 u(\mathbf{x}_{\mathrm{r}}, t; \mathbf{x}_{\mathrm{s}})}{\partial t^2} u_{\mathrm{res}}(\mathbf{x}_{\mathrm{r}}, t; \mathbf{x}_{\mathrm{s}}) \mathrm{d}t
$$
(13)

where u_{res} is the back propagated residual wavefield, and $v(\mathbf{x})$ is velocity.
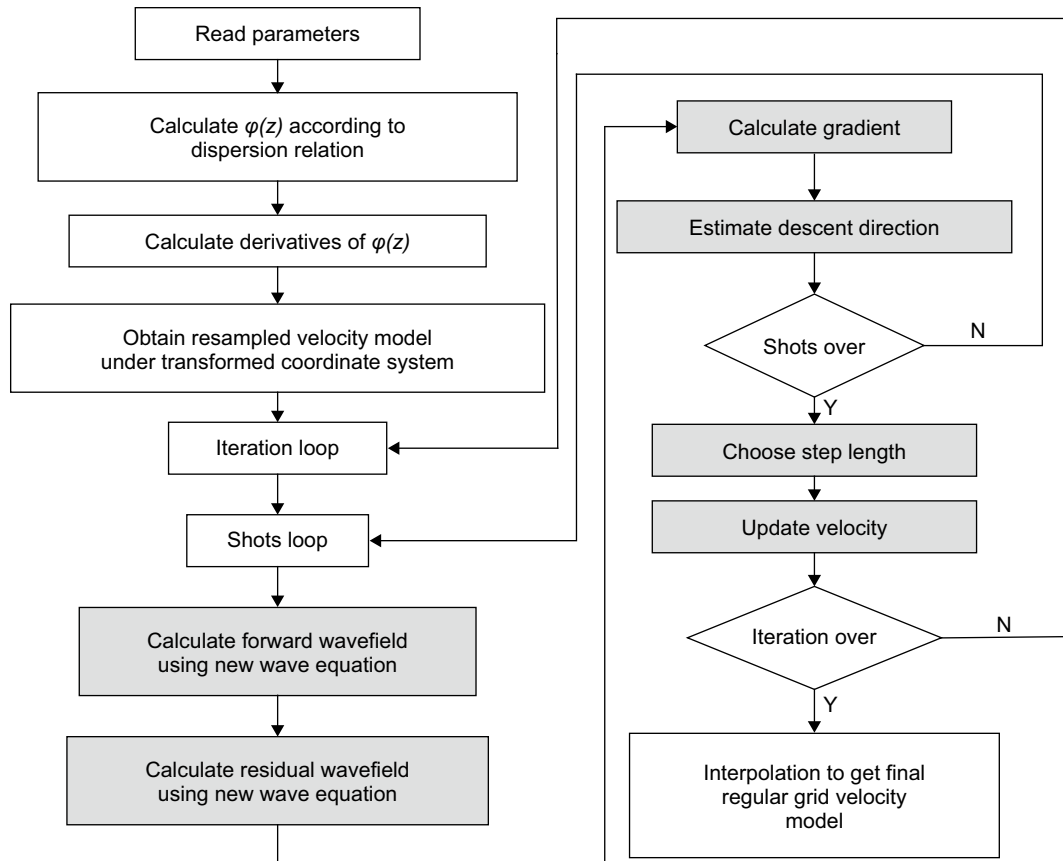
Fig. 3 Flowchart of adaptive variable-grid FWI

In this work, we use a nonlinear conjugate gradient method to determine the descent direction. The descent direction is defined as:

$$d_k = \begin{cases} -\nabla E(m_0), & k = 0 \\ -\nabla E(m_0) + \beta_k d_{k-1}, & k \geq 1 \end{cases} \quad (14)$$

where β_k is a correction coefficient modifying d_k into an optimal direction, and d_{k-1} is the descent direction in the previous iteration. Here, we use a hybrid scheme combining the Hestenes–Stiefel method and the Dai–Yuan method (Hager and Zhang 2006)

$$\beta_k = \max(0, \min(\beta_k^{HS}, \beta_k^{DY})) \quad (15)$$

where β_k^{HS} is calculated by the Hestenes–Stiefel method, and β_k^{DY} is estimated by the Dai–Yuan method.
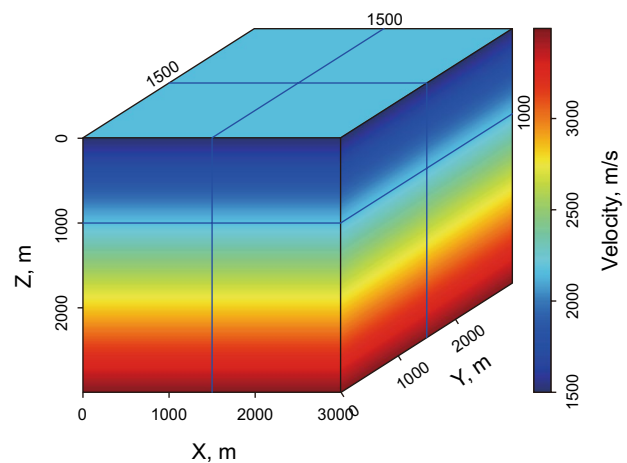
Fig. 4 Velocity model
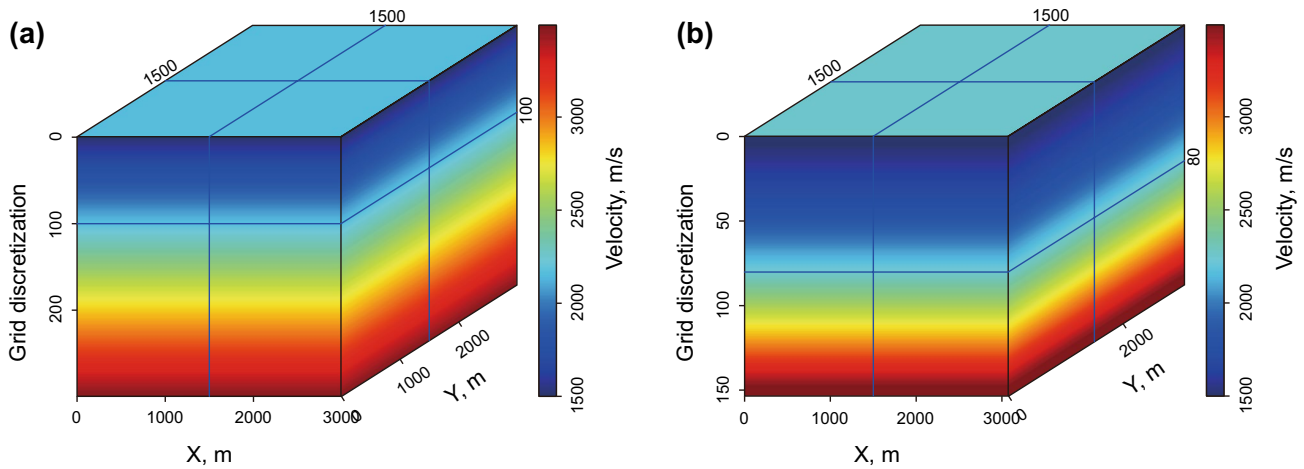
Fig. 5 Regular (**a**) and adaptive grid (**b**) discretization strategy

In this paper, we use the L_2 norm difference objective function minimized by a nonlinear conjugate gradient method based on the theory mentioned above. Based on a local minimization algorithm, the traditional FWI has a high risk of converging to local minima. To mitigate the "cycle skipping" problem caused by L_2 norm misfit, a multi-scale strategy is applied to lead the updated model to a better result.

3 3D numerical examples

3.1 Accuracy and efficiency tests

To validate the feasibility of the adaptive variable-grid algorithm, we designed 3D velocity models. Velocity changes with depth linearly from 1500 to 3500 m/s (Fig. 4). The model covers 3×3 km^2 laterally and is 3 km deep. The original regular-grid size is 10 m in three directions. We use the regular-grid finite-difference method and variable-grid finite-difference method to simulate wavefields, respectively. The sketches of both grid discretization strategies are shown in Fig. 5. It is shown that the grid number at depth decreases from 300 to 150 points, one-half of the original grid points. Besides, the velocity in deep layers is resampled using coarse grids, while near-surface layers are resampled using fine grids. The grid size increases from shallow layers to deep layers automatically. The source is a Ricker wavelet with a dominant frequency of 10 Hz. One source is added to the top surface in the middle.

We used both grid discretization methods to generate wavefields whose snapshot slices in three directions are extracted and shown in Fig. 6. The interpolated variable-grid wavefields are almost identical with regular-grid wavefields as can be seen by comparing Fig. 6a and b, c

and d, e and f. The difference between them is less than 1% of the regular-grid wavefields (Fig. 7), which validates that the new strategy of adaptive variable-grid discretization is accurate and credible.

According to the dispersion relation, the resampling interval depends on the velocity of each layer and seismic wavelet maximum frequency. To test the efficiency of the variable-grid finite-difference modeling method, we created several models with different ranges of velocities and record operation time of FWI based on both regular and variable-grid strategies. Table 2 presents some important modeling parameters. Among the series of tests, we treat test 1 as a standard, with 10-Hz dominant frequency, velocity ranging from 1500 m/s to 6000 m/s and 10-m original grid size. The computing time reduces to 50% of the regular-grid method. Therefore, the corresponding efficiency of the proposed variable-grid modeling method is twice as that of conventional modeling method. Then, the Ricker wavelet is set to 5 Hz and 20 Hz in tests 2 and 3 to find out the influence of the wavelet dominant frequency on efficiency improvement. In addition, the maximum velocity is changed to 3500 m/s in test 4, resulting in a slight decrease in efficiency improvement. The situation is similar when the original grid size is increased to 20 m in test 5. After several tests with different wavelet dominant frequencies, velocity range and original grid size, it is concluded that our adaptive variable-grid FWI displays obvious efficiency improvement compared with the conventional modeling method. In Fig. 8, we compare the normalized computing time of these tests, which shows that the computing time can decrease to less than 60% of the conventional method under most geological situations. Overall, the efficiency improvement ratio is about two in common geological conditions. Meanwhile, the accuracy loss is less than 1%, which can be completely tolerated if
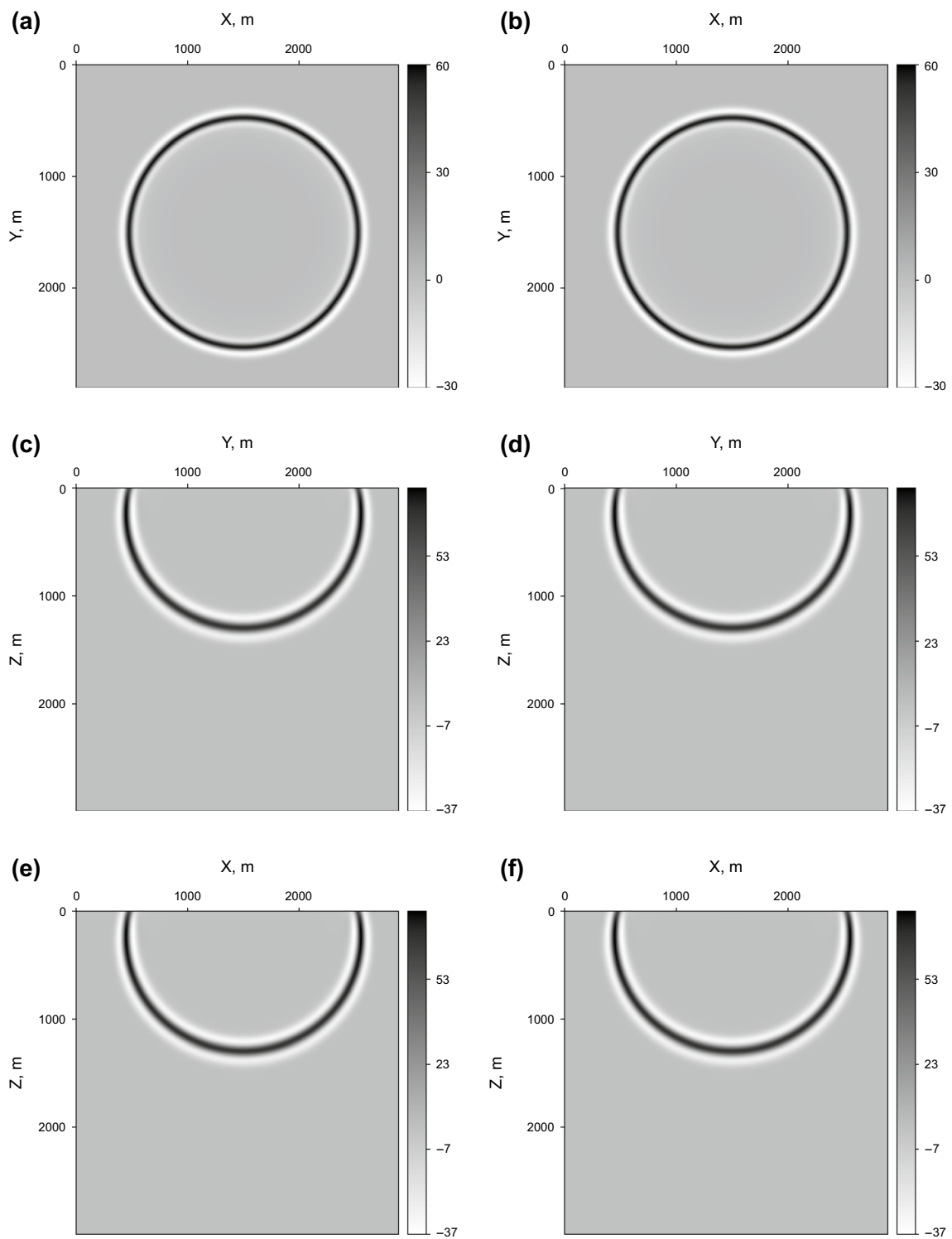
Fig. 6 Wavefield snapshot slices using regular-grid discretization at **a** top surface, **c** $x=1500$ m, **e** $y=1500$ m, and using variable-grid discretization at **b** top surface, **d** $x=1500$ m, **f** $y=1500$ m
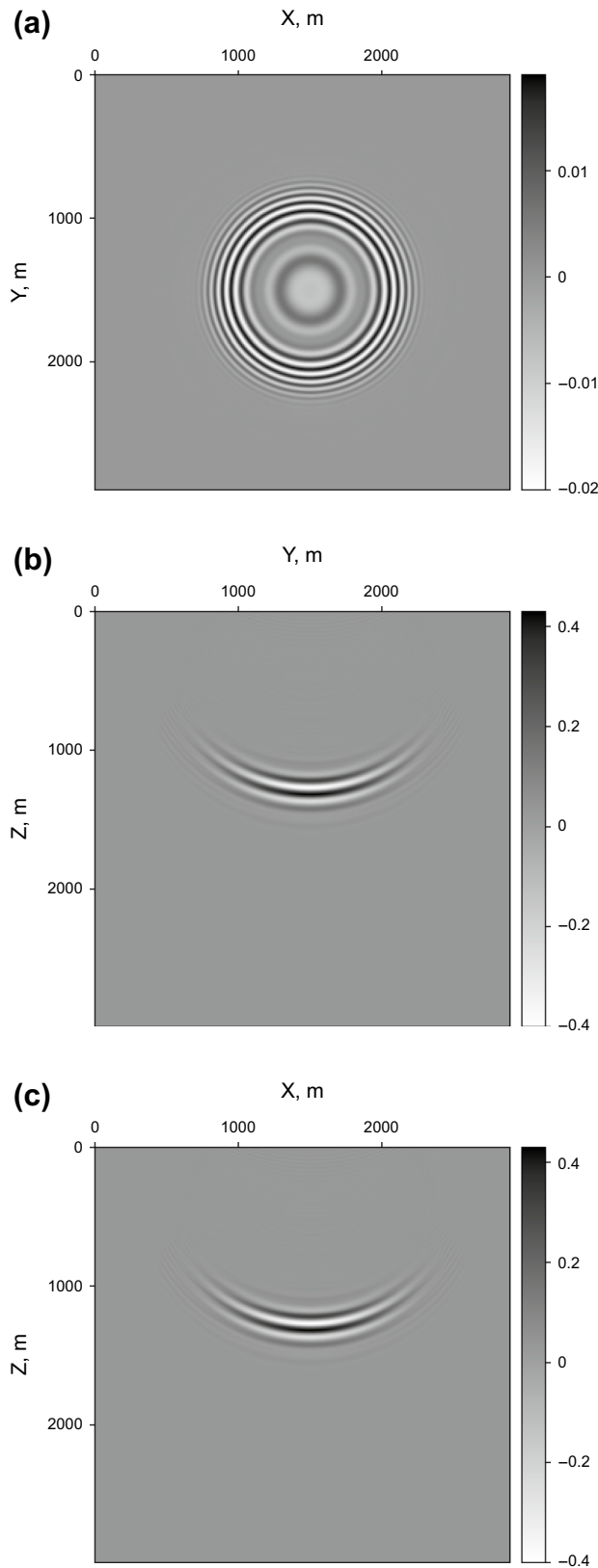
(a)

(b)

(c)

Fig. 7 The difference between wavefield snapshots using the two grid discretization strategy at **a** top surface, **b** $x = 1500$ m, **c** $y = 1500$ m

we apply the proposed accelerated modeling method to FWI.

3.2 3D application on the SEG/EAGE overthrust model

In this numerical test, a part of the 3D SEG/EAGE overthrust model is considered (Fig. 9). Due to limited available computer resources, our application is carried out on a shallow left part of the whole overthrust model. It is discretized on a $280 \times 45 \times 110$ grid with a grid spacing $h = 20$ m, which corresponds to a physical domain of 5.6 km $\times$ 0.9 km $\times$ 2.2 km. The minimum and maximum velocities in this overthrust model are 2000 m/s and 5500 m/s, respectively. We use fixed spread surface acquisition in which $28 \times 5 = 140$ sources are modeled every 200 m in both x and y directions. Receivers are regularly located each 20 m in the x direction and 40 m in the y direction. The acquisition geometry covers most of the underground area. A recording time of 2 s is used for the modeling. We use a time interval of 0.8 ms, resulting in 2500 time steps, to satisfy the stability condition.

The starting model for inversion is obtained by smoothing the true velocity model with a Gaussian function (Fig. 10). Slices of the exact model and initial model at constant $x = 2.2$ km and $y = 0.5$ km are presented. The constant depth slice at $z = 1$ km is chosen to show the lateral variation of the thrusted region as well as a channel complex, to demonstrate the horizontal accuracy of the 3D waveform inversion. The detailed channel information is removed by smoothing the exact model. In order to mitigate the local minima problem, the inversion is implemented using a multi-scale strategy, starting at the peak frequency of 3.5 Hz and sequentially stepping up to 7 Hz and then 15 Hz. For each scale, we compute 14 iterations.

The purpose of this experiment is to focus on the accuracy and efficiency of the proposed 3D variable-grid full-waveform inversion (VGFWI) compared with traditional full-waveform inversion. For this reason, the same acquisition geometry and workflow are implemented based on the regular-grid finite-difference modeling and the proposed variable-grid finite-difference modeling, respectively. After resampling the velocity model based on Eq. 3, we simulate the source wavefield and the residual wavefield using the new modified wave Eq. 10.

With a Ricker wavelet of 3.5-Hz dominant frequency, the velocity model is resampled to 54 grids in the depth direction, less than half of the original grid number (Table 3). The efficiency is double the traditional FWI with fixed square grids. Sequentially, the vertical resampled velocity model is decreased to 72 grids in depth, two-thirds of the original 110 grids, when the peak frequency is 7 Hz. However, as we invert for 13-Hz results, the new grid discretization strategy does not help to reduce computing time dramatically

Table 2 Efficiency comparison between variable-grid and conventional modeling

Test number	Original grid size, m	Wavelet dominant frequency, Hz	Velocity range, m/s	Normalized computing time, %	Efficiency improvement, %
1	10	10	1500–6000	50	200
2	10	5	1500–6000	34	300
3	10	20	1500–6000	59	171
4	10	10	1500–3500	59	171
5	20	10	1500–6000	58	173
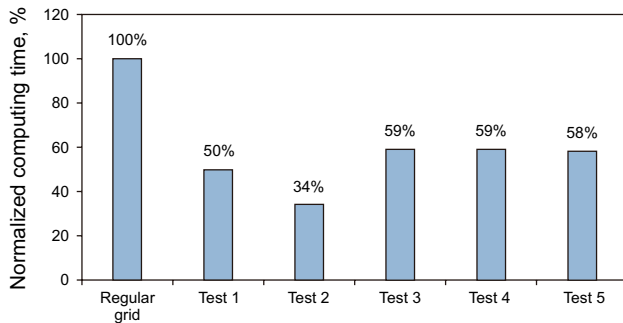
Fig. 8 Computing time comparison of variable-grid modeling and regular-grid modeling

Fig. 9 True velocity model (Horizontal slice locates at $z=1$ km. Inline section locates at $y=0.5$ km. Crossline section locates at $x=2.2$ km)

Fig. 10 Initial velocity model (Horizontal slice locates at $z=1$ km. Inline section locates at $y=0.5$ km. Crossline section locates at $x=2.2$ km)

to ensure the wavefield is adequately sampled so that dispersion never occurs. In Fig. 11, we compared the computing time of one iteration in particular, which shows that the overall computing time is reduced to less than two-thirds of the conventional FWI. Generally, the efficiency is almost twice that of traditional waveform inversion while updating background velocity with a low-dominant-frequency Ricker wavelet.

The inversion results using three-model scales are shown in Figs. 12 and 13, based on conventional FWI and variable-grid FWI (VGFWI), respectively. As shown in two sets of inverted velocity models, the large-scale layered structure

Table 3 Efficiency comparison between variable-grid and conventional FWI

	Dominant frequency of Ricker wavelet, Hz	Grid number, nz	Computing time of one iteration, s	Percentage of computing time, %	Efficiency improvement, %
Conventional FWI	–	110	9242	–	–
Variable-grid FWI	3.5	54	4452	48	208
	7	72	5651	61	163
	13	104	7629	82	121
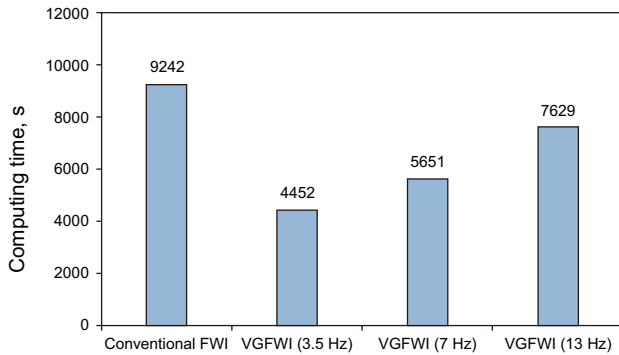
Fig. 11 Computing time comparison of variable-grid FWI and conventional FWI

is well characterized using both methods. From the inline section at $y=0.5$ km, it is observed that high-velocity layers inverted by VGFWI are obviously closer to true velocity compared with traditional FWI, which indicates that the proposed method converges faster. Depth slices at $z=1$ km show that variable-grid FWI also retains the accuracy of regular-grid FWI when reconstructing the target channel. The edges of channels are described with slightly higher resolution, and their velocities are retrieved closer to the exact model. Then, we extract vertical profiles from the shown inline section ($y=0.9$ km) at $x=1$ km and $x=2.7$ km (Fig. 14). The inverted curves using VGFWI are nearly identical with curves using FWI and true curves in the shallow region less than 1 km. Some deviations exist at depth due to relatively weak illumination and an insufficient number of iterations. Still, the whole inversion results of VGFWI achieve slightly higher agreement with true velocity than conventional FWI.

In general, the proposed adaptive variable-grid FWI is capable of introducing higher efficiency for waveform inversion. According to inversion results, such as sections and profiles, the accelerated FWI can achieve the same level of calculation precision as conventional FWI. If we compare inversion results more carefully, VGFWI shows slightly better agreement with the true model. Ha and Shin (2012) mentioned that the accumulated large-grid error could make the inversion using the transformed axis converge to a larger value. However, our method can control grid size in deep layers with milder z-axis variation according to the dispersion relation. Besides, the variable-grid method we proposed is free of dispersion. The grid size is sufficient for accurate modeling and inversion. That is the reason why the inversion results using the proposed method are not worse than conventional FWI. As for the slightly faster convergence rate, from our point of view, FWI is an inverse problem with nonlinearity and non-uniqueness, which lead to multiple solutions. The adaptive-grid velocity model has fewer grids especially for deep layers, so the inverse problem has less unknown variables mathematically. In this way, the complexity of the inverse problem is actually reduced, which makes it easier to converge.

In our method, only the vertical grid size changes because the main velocity change direction is with depth under common geological circumstances. However, when lateral change is dominant, multi-direction variable-grid modeling also needs to be developed to further improve efficiency. Another alternative development will be to try multiple grid time-space dual-variable methods, which will further decrease computation and involve multi-scale strategies by reducing grid scale through iterations.
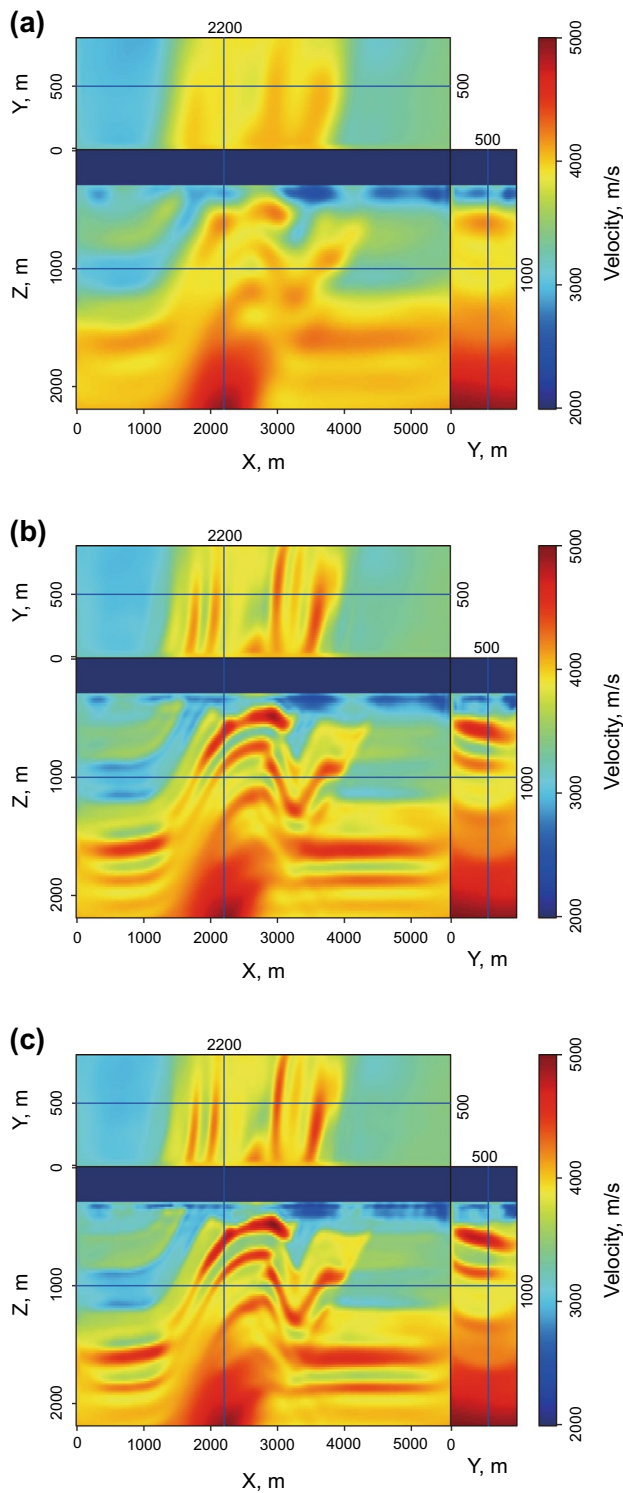
Fig. 12 Inverted velocity models using traditional FWI. **a** 14th iteration; **b** 28th iteration; **c** 42nd iteration (The horizontal slice is at $z = 1$ km. The inline section is at $y = 0.5$ km. The crossline section is at $x = 2.2$ km)
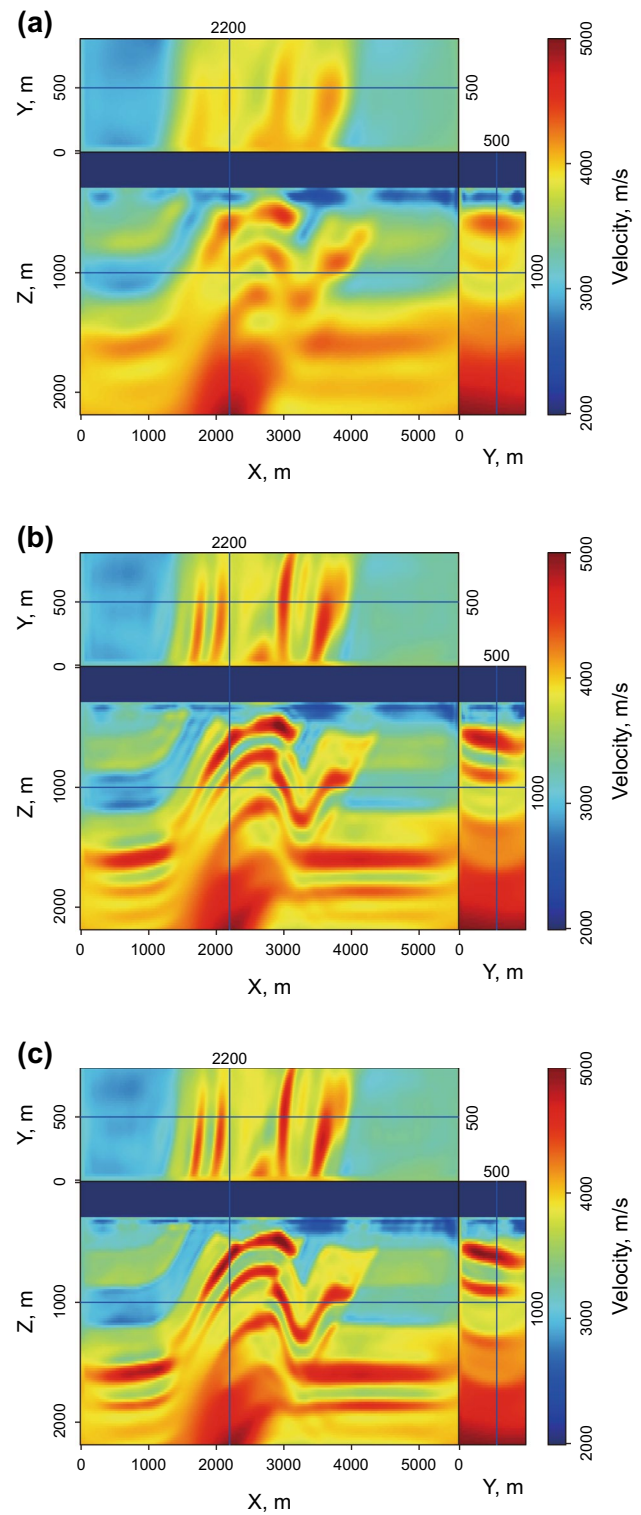
Fig. 13 Inverted velocity models using proposed variable-grid FWI. **a** 14th iteration; **b** 28th iteration; **c** 42nd iteration (The horizontal slice is at $z = 1$ km. The inline section is located at $y = 0.5$ km. The crossline section is at $x = 2.2$ km)
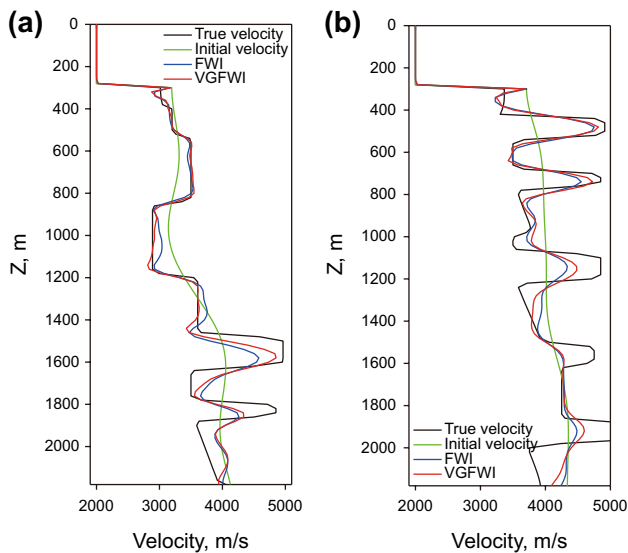
Fig. 14 Comparison between vertical profiles extracted from the true (black line), the starting (green line), conventional FWI (blue line), and variable-grid FWI (VGFWI) (red line). The three profile series are located at **a** $y = 0.9$ km, $x = 1$ km, **b** $y = 0.9$ km, $x = 2.7$ km

4 Conclusions

Full-waveform inversion in 3D cases is known to be limited by the problem of expensive computation when hundreds of shots and tens of iterations are required. The process of wavefield propagation is the most computationally intensive step in the implementation of FWI. To alleviate this problem, we proposed an adaptive variable-grid finite-difference modeling method by deriving a new modified acoustic wave equation. This grid discretization strategy is based on a dispersion relation, which improves modeling efficiency without losing accuracy. Then we apply this efficient wavefield simulation method to 3D full-waveform inversion on GPU devices. In this way, 3D FWI is dramatically accelerated without precision loss.

We have presented several applications on both simple and complex examples to give insight into the feasibility of our approach. Some preliminary tests on linear models have shown that the modeling method is adequately accurate with an error less than 1%. Moreover, the efficiency is increased to about twice that of the regular-grid finite-difference method under common geological conditions. After that, we have shown the 3D applications of variable-grid full-waveform inversion on part of the SEG/EAGE overthrust model. Owing to the proposed modeling method and CUDA-C language programming, the computing time is significantly reduced especially for large-scale background velocity inversion. In addition, cross sections, depth slices and profiles demonstrate that VGFWI is a cost-effective and quality-assured approach for 3D acoustic FWI.

References

Ben-Hadj-Ali H, Operto S, Virieux J. Velocity model building by 3D frequency-domain, full-waveform inversion of wide-aperture seismic data. Geophysics. 2008;73(5):VE101–17. https://doi.org/10.1190/1.2957948.

Ben-Hadj-Ali H, Operto S, Virieux J. An efficient frequency-domain full waveform inversion method using simultaneous encoded sources. Geophysics. 2011;76(4):R109–24. https://doi.org/10.1190/1.3581357.

Bunks C, Saleck FM, Zaleski S, Chavent G. Multiscale seismic waveform inversion. Geophysics. 1995;60(5):1457–73. https://doi.org/10.1190/1.1443880.

Clayton R, Engquist B. Absorbing boundary conditions for acoustic and elastic wave equations. Bull Seismol Soc Am. 1977;67(6):1529–40. https://doi.org/10.1007/BF02317997.

Fan N, Zhao LF, Gao YJ, Yao ZX. A discontinuous collocated-grid implementation for high-order finite-difference modeling discontinuous-grid FD modeling. Geophysics. 2015;80(4):T175–81. https://doi.org/10.1190/geo2015-0001.1.

Fan N, Zhao LF, Xie XB, Yao ZX. A discontinuous-grid finite-difference scheme for frequency-domain 2D scalar wave modeling. Geophysics. 2018;83(4):T235–44. https://doi.org/10.1190/geo2017-0535.1.

Fichtner A, Trampert J. Hessian kernels of seismic data functionals based upon adjoint techniques. Geophys J Int. 2011;185(2):775–98. https://doi.org/10.1111/j.1365-246X.2011.04966.x.

Ha W, Shin C. Efficient Laplace-domain modeling and inversion using an axis transformation technique. Geophysics. 2012;77(4):R141–8. https://doi.org/10.1190/geo2011-0424.1.

Hager WW, Zhang H. A survey of nonlinear conjugate gradient methods. Pac J Optim. 2006;2(1):35–58. https://doi.org/10.1006/jsco.1995.1040.

Huang JP, Qu YM, Li QY, Li ZC, Li GL, Bu CC, et al. Variable-coordinate forward modeling of irregular surface based on dual-variable grid. Appl Geophys. 2015;12(1):101–10. https://doi.org/10.1007/s11770-014-0476-2.

Jastram C, Behle A. Acoustic modelling on a grid of vertically varying spacing. Geophys Prospect. 1992;40(2):157–69. https://doi.org/10.1111/j.1365-2478.1992.tb00369.x.

Jastram C, Tessmer E. Elastic modelling on a grid with vertically varying spacing. Geophys Prospect. 1994;42(4):357–70. https://doi.org/10.1111/j.1365-2478.1994.tb00215.x.

Jiang J, Zhu P. Acceleration for 2D time-domain elastic full waveform inversion using a single GPU card. J Appl Geophys. 2018;152:173–87. https://doi.org/10.1016/j.jappgeo.2018.02.015.

Li Y, Li Z, Zhang K. Frequency-domain waveform inversion with irregular surface based on variable grid finite difference method. In: SPG/SEG 2016 international geophysical conference. 2016; p. 305–7. https://doi.org/10.1190/igcbeijing2016-088.

Li Z, Li Q, Huang J, Na L, Kun T. A stable and high-precision dual-variable grid forward modeling and reverse time migration method.

Geophys Prospect Pet. 2014;53(2):127–36. https://doi.org/10.3969/j.issn.1000-1441.2014.02.001 **(in Chinese)**.

Liu L, Ding R, Liu H, Liu H. 3D hybrid-domain full waveform inversion on GPU. Comput Geosci. 2015;83:27–36. https://doi.org/10.1016/j.cageo.2015.06.017.

Ma Y, Hale D. Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion. Geophysics. 2013;78(6):R223–33. https://doi.org/10.1190/geo2013-0004.1.

Métivier L, Brossier R, Mérigot Q, Oudet E, Virieux J. Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion. Geophys J Int. 2016;205(1):332–64. https://doi.org/10.1093/gji/ggw014.

Moczo P. Finite-difference technique for SH-waves in 2-D media using irregular grids—application to the seismic response problem. Geophys J Int. 1989;99(2):321–9. https://doi.org/10.1111/j.1365-246X.1989.tb01691.x.

Plessix R.-E. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. Geophy J Int. 2006;167(2):495–503.

Plessix RE, Perkins C. Thematic set: full waveform inversion of a deep water ocean bottom seismometer dataset. First Break. 2010;28(1728):71–8. https://doi.org/10.3997/1365-2397.2010013.

Qu Y, Li Z, Huang J, Li Q, Li J. Multiple dual-variable grid encoding full time inversion based on an optimized encoding function. In: Workshop: depth model building: full-waveform inversion. 2015; p. 125–9. https://doi.org/10.1190/fwi2015-031.

Raknes EB, Arntsen B, Weibull W. Three-dimensional elastic full waveform inversion using seismic data from the Sleipner area. Geophys J Int. 2015;202(3):1877–94. https://doi.org/10.1093/gji/ggv258.

Shin C, Cha YH. Waveform inversion in the Laplace domain. Geophys J Int. 2008;173(3):922–31. https://doi.org/10.1111/j.1365-246X.2008.03768.x.

Shin J, Ha W, Jun H, Min DJ, Shin C. 3D Laplace-domain full waveform inversion using a single GPU card. Comput Geosci. 2014;67:1–13. https://doi.org/10.1016/j.cageo.2014.02.006.

Sirgue L, Etgen JT, Albertin U. 3D frequency domain waveform inversion using time domain finite difference methods. In: 70th EAGE conference and exhibition incorporating SPE EUROPEC 2008. https://doi.org/10.3997/2214-4609.20147683.

Sun C, Li S, Ni C. Wave equation numerical modeling by finite difference method with varying grid spacing. Geophys Prospect Pet. 2008;47(2):123–7. https://doi.org/10.1002/clen.200700058.

Tarantola A. Inversion of seismic reflection data in the acoustic approximation. Geophysics. 1984;49(8):1259–66. https://doi.org/10.1190/1.1441754.

Van Leeuwen T, Mulder WA. A correlation-based misfit criterion for wave-equation traveltime tomography. Geophys J Int. 2010;182(3):1383–94. https://doi.org/10.1111/j.1365-246X.2010.04681.x.

Vigh D, Starr EW. 3D prestack plane-wave, full-waveform inversion. Geophysics. 2008;73(5):VE135–44. https://doi.org/10.1190/1.2952623.

Virieux J, Operto S. An overview of full-waveform inversion in exploration geophysics. Geophysics. 2009;74(6):WCC1–26. https://doi.org/10.1190/1.3238367.

Warner M, Ratcliffe A, Nangoo T, Morgan J, Umpleby A, Shah N, et al. Anisotropic 3D full-waveform inversion. Geophysics. 2013;78(2):R59–80. https://doi.org/10.1190/geo2012-0338.1.

Weiss RM, Shragge J. Solving 3D anisotropic elastic wave equations on parallel GPU devices. Geophysics. 2013;78(2):F7–15. https://doi.org/10.1190/geo2012-0063.1.

Yang P, Gao J, Wang B. A graphics processing unit implementation of time-domain full-waveform inversion. Geophysics. 2015;80(3):F31–9. https://doi.org/10.1190/geo2014-0283.1.

Zhang K, Ma X, Li Y, Wu H, Cui C, Zhang X, et al. Parameter prediction of hydraulic fracture for tight reservoir based on micro-seismic and history matching. Fractals. 2018;26(02):1840009. https://doi.org/10.1142/S0218348X18400091.

Zhu SW, Qu SL, Wei XC, Liu CY. Numeric simulation by grid-various finite-difference elastic wave equation. Oil Geophys Prospect. 2007;42(6):634–9. https://doi.org/10.1016/S1872-5813(07)60034-6 **(in Chinese)**.