Research Article

Preconditioned Conjugate Gradient Algorithm-Based 2D Waveform Inversion for Shallow-Surface Site Characterization

Jianbo Guan ⓘ**, Yu Li** ⓘ**, and Guohua Liu** ⓘ

School of Geological Engineering and Geomatics, Chang'an University, Xi'an, 710054, China

Correspondence should be addressed to Yu Li; liyupa@chd.edu.cn

The full-waveform inversion (FWI) of a Love wave has become a powerful tool for shallow-surface site characterization. In classic conjugate gradient algorithm- (CG) based FWI, the energy distribution of the gradient calculated with the adjoint state method does not scale with increasing depth, resulting in diminished illumination capability and insufficient model updating. The inverse Hessian matrix (HM) can be used as a preprocessing operator to balance, filter, and regularize the gradient to strengthen the model illumination capabilities at depth and improve the inversion accuracy. However, the explicit calculation of the HM is unacceptable due to its large dimension in FWI. In this paper, we present a new method for obtaining the inverse HM of the Love wave FWI by referring to HM determination in inverse scattering theory to achieve a preconditioned gradient, and the preconditioned CG (PCG) is developed. This method uses the Love wave wavefield stress components to construct a pseudo-HM to avoid the huge calculation cost. It can effectively alleviate the influence of nonuniform coverage from source to receiver, including double scattering, transmission, and geometric diffusion, thus improving the inversion result. The superiority of the proposed algorithm is verified with two synthetic tests. The inversion results indicate that the PCG significantly improves the imaging accuracy of deep media, accelerates the convergence rate, and has strong antinoise ability, which can be attributed to the use of the pseudo-HM.

1. Introduction

The shear modulus is one of the vital engineering parameters directly related to the strength of a near-surface medium, with the shear wave (S-wave) velocity being a critical indicator. The multichannel analysis of surface waves based on dispersion curves has been one of the most accepted methods of shallow-surface S-wave velocity imaging [1, 2]. However, this method applies to a pseudo-layered medium that has gentle horizontal variations, and the resultant image of a pseudo-two-dimensional (2D) profile is formed by interpolation. This type of procedure limits the vertical and horizontal resolution of the method. Therefore, a method for extracting dispersion curve clusters by scanning multichannel seismic data with the application of the localized slant stack method to multichannel single-shot records has been presented to achieve 1.5D imaging [3]. This method, however, takes advantage of the multichannel dispersion information to the maximum extent and still cannot be used

to implement 2D or 3D imaging. Furthermore, Li et al. [4, 5] proposed a dispersion curve wave-equation theory that formulated wave-equation dispersion (WD) inversion. 2D S-wave velocity imaging could be directly implemented in this method, and the method was free from the cycle-skipping phenomenon. However, this method required preprocessing data via mode decomposition and noise suppression, so the ground-state energy in the surface waves was dominant overall frequency bands and not perturbed by the excited-state energy. These procedures are complicated, and the images acquired do not have high precision. For the above methods based on dispersion analysis, it is difficult to obtain quality results in shallow exploration, and the engineering objectives sometimes cannot be met.

By using all the amplitude and phase information for seismic datasets, full-waveform inversion (FWI) [6, 7] matches the observed and predicted data to reveal a high-resolution and high-precision underground velocity model, and it can meet the need for velocity parameters imposed by

complex structural imaging. This technique has gradually become a research hotspot in the field of seismic data inversion [8–10]. Previous FWI works have primarily focused on the inversion of P and S waves [11–14], but the development of surface wave FWI has become feasible [8] due to the dominance of surface wave energy in the near-surface wavefield (the vertically induced Rayleigh wave accounts for 70% of the horizontally induced Love wave for 90% of the near-surface wavefield) [15]. Compared with a Rayleigh wave, a Love wave is formed by the interference of multi-reflected SH-waves near a free surface [16] and is independent of the P-wave velocity while being more sensitive to the S-wave velocity. These properties have led to a relatively efficient and straightforward FWI framework for Love waves. Gradient-based local optimization algorithms such as the steepest descent algorithm [6] and conjugate gradient algorithm [17] have been widely used in solving FWI, a large-scale and strong nonlinear optimization problem, with low computational cost. However, the calculated gradient of the adjoint-state method [18] contains a singular value and uneven energy distribution, both of which are caused by the double scattering and limitations in the wavefield illumination and the wavelet frequency band. Hence, the model gradient amplitude at depth decreases rapidly, along with the diminished illumination capability, and the model parameters cannot be updated with apparent improvements.

To alleviate this problem, four different approaches have been proposed and developed: the distance-weighting, layer-stripping, and full Hessian matrix (HM) schemes and the approximate-HM strategies. The distance-weighting [19, 20] method adjusts the gradient amplitude across the model space by multiplying the gradient by a function of the distance from the source, and this method balances the energy distribution of the gradient by giving greater weight to the far offset and deep regions. Based on this idea, the layer-stripping method [21] divides a model into several parts according to different depths, and these parts are updated continuously according to the rules, from shallow to deep. Therefore, the gradient in the deep region is not suppressed due to the large gradient amplitude in the shallow region, and the reconstruction effect in the deep region is improved. Both methods can improve the accuracy of model reconstruction to a certain extent. However, the function of distance or depth cannot completely or precisely express the scaling of a gradient. There is no objective standard to ensure the concrete implementation of these methods. For example, there are no standards for determining how to set the weight value of distance or depth in the distance-weighting method or how to determine where and when to move from one layer to the next in the layer-stripping method. Different criteria will lead to different inversion results. The full-HM method relies on the advantages of second-order optimization algorithms such as the Newton [7] and Gauss-Newton [22] algorithms. A second-order partial derivative of the cost function, known as the HM, contains information about the curvature that can be used to clearly predict the illumination deficiency at depth in the model [14]. Thus, an inverse HM can adjust the ratio and balance the correction terms of the gradient to strengthen the illumination of the deep part of the model. Pratt et al. [22] implemented FWI in the frequency-spatial domain through a Full-Newton algorithm and provided a complete calculation method of the full-HM. This method greatly improved the inversion accuracy, but the calculation cost was too expensive to be acceptable under conventional conditions. In addition, Métivier et al. [14] calculated the full-HM based on the second-order adjoint state method and proposed the truncated-Newton method, which improved the calculation efficiency and ensured the robustness compared with the Full-Newton method. After that, some new Newton-class algorithms emerged [23, 24], but none of them could greatly reduce the computational cost of the full-HM. The approximate-HM employs the idea of estimation or construction to obtain the HM, solving the dilemma of determining the HM explicitly with current computing capabilities. The most common method is calculating the HM with the reciprocity theorem and the virtual source theory. Pratt et al. [22] pointed out that the approximate-HM was a diagonally dominant matrix with a high-frequency approximation. Only M diagonal elements in the HM are needed in the inversion involving M parameters. This fact greatly simplifies the calculation of the HM. Shin et al. [25] provided an estimating method for the diagonal terms in the approximate-HM in prestack depth migration based on inverse scattering theory. However, they failed to complete factors related to the geometric spreading. Later, Choi et al. [26] improved this estimation method by considering the amplitude of the seismic data in calculating the diagonal terms. However, this new approximate-HM only explained the source geometric spreading. Another approach is to construct the approximate-HM by combining information from a gradient, model, and cost function, including L-BFGS [27, 28], the corrected pseudo-Newton algorithm [10], and the pseudo-Newton algorithm [29]. These methods have further simplified the approximation method of the HM and achieved relatively effective results, but some accuracy has been lost.

In this paper, we present a new method to estimate the HM in Love wave FWI based on the calculation method of the HM in inverse scattering theory, and the preconditioned CG (PCG) is developed in combination with the CG. The proposed method uses the stress components of the Love wave wavefield to construct a pseudo-inverse HM to filter, balance, and regularize the gradient and alleviate the influence of nonuniform coverage from source to receiver, including double scattering, transmission, and geometric diffusion, thus improving the inversion accuracy. Because the required stress components have been obtained in the forward simulation, no extra calculation overhead is needed, thus ensuring the calculation efficiency and saving the calculation costs. Moreover, the construction of the gradient and the HM are derived in detail, and the implementation procedure of the PCG-based Love wave FWI is described in detail. The trial results from the checkerboard model and the complex structure model indicate that compared with the classical CG, the PCG can clearly rebuild the stratigraphic anomaly boundary and reconstruct the primary stratigraphic anomaly at depth, and it has strong antinoise ability. The advantages of PCG are due to the presence of the pseudo-HM.

2. Methodology

The core concepts of FWI are to fit the predicted wavefield from the forwarding modeling to that observed in the field and minimize the misfit between these two datasets. Therefore, the optimal modeled data can be achieved. The optimization includes several key steps, such as the forward simulation of the seismic wavefield, the establishment of the optimization algorithm, the determination of the gradient, and the determination of the HM of a cost function.

2.1. Forward Problems. Wave equations describe the physical process of seismic waves as they travel through a medium. The forward problem involves obtaining the wavefield at the receiving point by solving the wave equation when the underground medium parameters and the source function are known. In a 2D isotropic medium, assuming that a particle has only a nonzero displacement in the y-direction of the X-Z plane (X is the horizontal distance and Z is the depth), coupled with Hooke's Law, the pseudo-conservative Love wave elastodynamic equation [30] in the time domain is

$$
\begin{cases}
\rho(x,z)\dfrac{\partial v_y(x,z,t)}{\partial t} = \dfrac{\partial \sigma_{xy}(x,z,t)}{\partial x} + \dfrac{\partial \sigma_{zy}(x,z,t)}{\partial z} + F(x,z,t), \\[2mm]
\dfrac{\partial \sigma_{xy}(x,z,t)}{\partial t} = \rho(x,z)V_s^2(x,z)\dfrac{\partial v_y(x,z,t)}{\partial x}, \\[2mm]
\dfrac{\partial \sigma_{zy}(x,z,t)}{\partial t} = \rho(x,z)V_s^2(x,z)\dfrac{\partial v_y(x,z,t)}{\partial z},
\end{cases}
\tag{1}
$$

where $\rho(x,z)$ is the density, $V_s(x,z)$ is the S-wave velocity, $\sigma_{xy}(x,z,t)$ and $\sigma_{zy}(x,z,t)$ are the shear stresses, $v_y(x,z,t)$ is the particle velocity, and $F(x,z,t)$ is the seismic source excited by an external force. A general form of (1) is

$$
D(\mathbf{m})\mathbf{u}(\mathbf{m}) = \varphi,
\tag{2}
$$

where $D(\mathbf{m})$ is the partial-differential operator, $\mathbf{u}(\mathbf{m})$ is the seismic wavefield, φ is the external force loaded, $\mathbf{m}$ is the parameterized model in terms of the S-wave velocity and the density, $\mathbf{m} = (\mathbf{V}_s, \text{ove it rem}) = (m_1, m_2, \ldots, m_N)$, and N is the number of model parameters. The high-order staggered-grid finite-difference method [30] with second-order temporal accuracy and tenth-order spatial accuracy is used to solve (1) and (2) discretely. The free-surface boundary of the top part of the model is processed with the stress-image method, and the bottom, left, and right boundaries of the model are processed with the multiaxial perfectly matched layer technique. The number of absorption layers in three directions is 50. The source is a Ricker wavelet, and it is loaded into $v_y(x,z,t)$.

2.2. Inversion Problems. FWI is a highly nonlinear problem, so nonlinear inversion theory should be adopted to solve it. In order to measure the fitting degree of the predicted data $\mathbf{d}_{\text{pre}}$ and the observed data $\mathbf{d}_{\text{obs}}$, as well as the residual energy

between these data, we define an L2-norm-based cost function:

$$
E(\mathbf{m}) = \frac{1}{2}\left\| \mathbf{d}_{\text{pre}}(\mathbf{m}) - \mathbf{d}_{\text{obs}} \right\|^2 = \frac{1}{2}\left\| R\mathbf{u}(\mathbf{m}) - \mathbf{d}_{\text{obs}} \right\|^2,
\tag{3}
$$

where $\| \cdot \|^2$ denotes the L2-norm calculation, R is the mapping from the wavefield to the receiver arrays, and $\mathbf{u}(\mathbf{m})$ is the solution to equation (2). The inversion process involves finding a parameterized model that perfectly predicts the observed data by minimizing the cost function through a global or local optimization method. The global algorithm can basically avoid the local minimum problem, but a large amount of computation hinders its popularization and application. The local algorithm [31] is still widely used at present, and it is implemented by updating the model parameters iteratively, for which the iteration sequence is expressed as

$$
\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k,
\tag{4}
$$

where $\mathbf{m}_k$, α_k, and $\Delta\mathbf{m}_k$ are the model parameters of the kth iteration, the optimal step length estimated with the parabolic fitting method [32], and the descent direction of the model update, respectively. A Taylor series expansion with second-order precision is performed on the cost function (3) at the $(k+1)$th iteration:

$$
E(\mathbf{m}_{k+1}) = E(\mathbf{m}_k) + \Delta\mathbf{m}_k\left(\frac{\partial E(\mathbf{m}_k)}{\partial \mathbf{m}_k}\right)
$$
$$
+ \frac{1}{2}\Delta\mathbf{m}_k\left(\frac{\partial^2 E(\mathbf{m}_k)}{\partial^2 \mathbf{m}_k}\right)\Delta\mathbf{m}_k^T + \mathrm{O}\left(\left\|\Delta\mathbf{m}_k^3\right\|\right).
\tag{5}
$$

The derivative of the cost function $E(\mathbf{m}_{k+1})$ with respect to $\Delta\mathbf{m}_k$ is assumed to be zero in order to find the minimum of the cost function near the model $\mathbf{m}_k$:

$$
\frac{\partial E(\mathbf{m}_{k+1})}{\partial \Delta\mathbf{m}_k} = \frac{\partial E(\mathbf{m}_k)}{\partial \mathbf{m}_k} + \Delta\mathbf{m}_k\left(\frac{\partial^2 E(\mathbf{m}_k)}{\partial^2 \mathbf{m}_k}\right) = 0.
\tag{6}
$$

The descent direction is obtained by rearranging (6):

$$
\Delta\mathbf{m}_k = -\left(\frac{\partial^2 E(\mathbf{m}_k)}{\partial^2 \mathbf{m}_k}\right)^{-1}\frac{\partial E(\mathbf{m}_k)}{\partial \mathbf{m}_k} = -\mathbf{H}_k^{-1}\nabla E(\mathbf{m}_k),
\tag{7}
$$

where $\nabla E(\mathbf{m}_k)$ and $\mathbf{H}_k^{-1}$ are the gradient and the inverse HM of the cost function, respectively.

The inverse HM in the gradient-based local algorithms is substituted for by the identity matrix, and the CG is widely used in FWI with the descent direction:

$$
\begin{cases}
\Delta\mathbf{m}_0 = -\nabla E(\mathbf{m}_0), \\[2mm]
\Delta\mathbf{m}_{k+1} = -\nabla E(\mathbf{m}_{k+1}) + \beta_{k+1}\Delta\mathbf{m}_k, \\[2mm]
\beta_{k+1} = \dfrac{\nabla E(\mathbf{m}_{k+1})^T\left(\nabla E(\mathbf{m}_{k+1}) - \nabla E(\mathbf{m}_k)\right)}{\nabla E(\mathbf{m}_k)^T\nabla E(\mathbf{m}_k)},
\end{cases}
\tag{8}
$$

where β_{k+1} is the scalar parameter of the $(k+1)$th iteration and the superscript T denotes the matrix transpose operator.

The core idea of the preconditioned gradient-based local algorithms is to solve or construct the inverse HM (7), and the descent direction is

$$
\begin{cases}
\Delta\mathbf{m}_0 = -\mathbf{H}_0^{-1}\nabla E(\mathbf{m}_0), \\
\Delta\mathbf{m}_{k+1} = -\mathbf{H}_{k+1}^{-1}\nabla E(\mathbf{m}_{k+1}) + \beta_{k+1}^P\Delta\mathbf{m}_k,
\end{cases}
$$

$$
\beta_{k+1}^P = \frac{\left(\mathbf{H}_{k+1}^{-1}\nabla E(\mathbf{m}_{k+1})\right)^T\left(\mathbf{H}_{k+1}^{-1}\nabla E(\mathbf{m}_{k+1}) - \mathbf{H}_k^{-1}\nabla E(\mathbf{m}_k)\right)}{\left(\mathbf{H}_k^{-1}\nabla E(\mathbf{m}_k)\right)^T\left(\mathbf{H}_k^{-1}\nabla E(\mathbf{m}_k)\right)}.
$$

$$(9)$$

The value of β^P is calculated with a formula from Métivier et al. [14]. To accelerate the convergence of the cost function, the scalar parameter is made nonnegative [33]:

$$
\beta_{k+1} = \max\{0, \beta_{k+1}\}. \tag{10}
$$

2.3. Adjoint-State Gradient and Preconditioned Operator. The acquisition of the cost function gradient is the core of the local optimization methods, which can be calculated with the explicit method [8] and the adjoint state method [18]. The explicit method obtains gradients by calculating the actual derivatives of the cost function concerning the various model parameters, which requires as much forward calculation as the model parameters. This type of computation is unacceptable for larger models. The adjoint state method

overcomes this problem and only requires two forward calculations to obtain the gradient. In the adjoint-state method, it is pointed out that the gradient is the zero-lag cross-correlation of the forward-propagating wavefield and the backward-propagating wavefield obtained from the backward-propagating data residual at the geophone. This relationship is presented with detailed derivation in [34]. The backward-propagating wavefield can be determined with the adjoint function as follows:

$$
D(\mathbf{m})^{*H}\gamma = -R^{*H}(\mathbf{d}_{\mathrm{obs}} - R\mathbf{u}(\mathbf{m})), \tag{11}
$$

where $(\cdot)^{*H}$ denotes the conjugate transpose, γ is the backward-propagating wavefield to be calculated, and $\mathbf{u}(m)$ is the solution to (2) corresponding to the designated seismic source terms. Specifically, the source term on the right-hand side of (2) is a reverse time series consisting of the differences between the observed data and the predicted seismic data. Hence, the cost function gradient to the model parameters can be calculated according to the following equation:

$$
\nabla E(\mathbf{m}) = \mathrm{real}\left(\partial_{\mathbf{m}}D(\mathbf{m})\mathbf{u}(\mathbf{m}), \gamma(\mathbf{m})\right)_\omega, \tag{12}
$$

where real denotes real arithmetic and $(\cdot,\cdot)_\omega$ is the scalar product of the enclosed terms in the inner product space ω. Based on this equation, combined with the chain rule [17], the Love wave FWI gradient can be calculated with respect to the S-wave velocity and the density:

$$
\nabla E_{\mathrm{vel}}(\mathbf{m}) = -\frac{2}{\rho\mathbf{V}_s^3}\int_1^S\int_0^T\left[\frac{\partial\boldsymbol{\sigma}_{xy}^f}{\partial t}\boldsymbol{\sigma}_{xy}^b + \frac{\partial\boldsymbol{\sigma}_{zy}^f}{\partial t}\boldsymbol{\sigma}_{zy}^b\right]\mathrm{d}t\mathrm{d}s, \tag{13}
$$

$$
\nabla E_{\mathrm{den}}(\mathbf{m}) = -\int_1^S\int_0^T\left[\frac{\partial\mathbf{v}_y^f}{\partial t}\mathbf{v}_y^b + \frac{1}{\rho^2\mathbf{V}_s^2}\left(\frac{\partial\boldsymbol{\sigma}_{xy}^f}{\partial t}\boldsymbol{\sigma}_{xy}^b + \frac{\partial\boldsymbol{\sigma}_{zy}^f}{\partial t}\boldsymbol{\sigma}_{zy}^b\right)\right]\mathrm{d}t\mathrm{d}s, \tag{14}
$$

where ∇E_{vel} and ∇E_{den} are the gradients of the S-wave velocity and the density, respectively, S represents the number of shots, T represents the sampling points in the time domain, σ_{xy}^f and σ_{zy}^f are the shear stresses of the forward-propagating wavefields, and σ_{xy}^b and σ_{zy}^b are the shear stresses of the backward-propagating wavefields.

Shin et al. [25] promoted the development of the preconditioned approach in FWI. Due to the narrow frequency band of the observed data, the double scattering, and other reasons, the gradient amplitude calculated by (13) and (14) does not scale with the depth increase. Therefore, the illumination with the depth is lost, and the parameters related to the medium cannot be updated. The HM can clearly predict the above problems because it contains the curvature information. The inverse HM, also known as the preconditioned operator, can be used as a deconvolution operator to adjust the proportion and balance the correction terms of the gradient, eliminate the

limited band effect, and enhance illumination for deep regions. However, the vast dimensions of the inverse HM make it almost impossible to calculate this matrix with the current computing power. Nevertheless, we can construct a pseudo-HM using the reciprocity theorem [14, 25]. In this study, a new pseudo-HM in Love wave FWI is solved by referring to the HM calculation method of the inverse scattering theory [25], which can suppress the artifacts caused by multiple scattering in the gradient. Based on this, the form of the pseudo-HM can be expressed as

$$
\mathbf{H}_{ij}(\mathbf{m}) = \left(\left[\partial_{\mathbf{m}_i}D(\mathbf{m})\mathbf{u}(\mathbf{m})\right]^T\left[\partial_{\mathbf{m}_j}D(\mathbf{m})\mathbf{u}(\mathbf{m})\right]\right);
$$
$$
i, j = 1, \ldots, \mathrm{M}, \tag{15}
$$

where M is the model parameter space. This equation is transformed into a parameterized expression of the S-wave velocity and the density:

$$\mathbf{H}_{\mathrm{vel}_{ij}} = \frac{4}{\left(\boldsymbol{\rho}\mathbf{V}_s^3\right)^2} \int_1^S \int_0^T \left[\frac{\partial\boldsymbol{\sigma}_{xyi}^f}{\partial t}\frac{\partial\boldsymbol{\sigma}_{xyj}^f}{\partial t} + \frac{\partial\boldsymbol{\sigma}_{zyi}^f}{\partial t}\frac{\partial\boldsymbol{\sigma}_{zyj}^f}{\partial t}\right]\mathrm{d}t\mathrm{d}s, \tag{16}$$

$$\mathbf{H}_{\mathrm{den}_{ij}} = \int_1^S \int_0^T \left[\frac{\partial\mathbf{v}_{yi}^f}{\partial t}\frac{\partial\mathbf{v}_{yj}^f}{\partial t} + \frac{1}{\left(\boldsymbol{\rho}^2\mathbf{V}_s^2\right)^2}\left(\frac{\partial\boldsymbol{\sigma}_{xyi}^f}{\partial t}\frac{\boldsymbol{\sigma}_{xyj}^f}{\partial t} + \frac{\partial\boldsymbol{\sigma}_{zyi}^f}{\partial t}\frac{\partial\boldsymbol{\sigma}_{zyj}^f}{\partial t}\right)\right]\mathrm{d}t\mathrm{d}s, \tag{17}$$

where $\mathbf{H}_{\mathrm{vel}}$ and $\mathbf{H}_{\mathrm{den}}$ are the pseudo-HMs of the S-wave velocity and the density, respectively. The inverse HM is diagonally dominant for the high-frequency approximation [22], and only the diagonal elements must be calculated. In addition, an adjustment factor is added to avoid the inversion instability caused by the near-zero value generated by the geometric diffusion of the wavefield on the diagonal. The preconditioned operator is adjusted at the kth iteration to

$$\mathbf{H}_k^{-1} = \mathrm{diag}\left(\frac{1}{\mathbf{H}_{ii_k} + \varepsilon\mathrm{maximize}\left(\mathbf{H}_{ii_k}\right)}\right), \tag{18}$$

where ε is an empirical parameter. The algorithm has high performance when $\varepsilon = 10^{-5} - 10^{-2}$ after various numerical tests. Eventually, the preconditioned operator is adjusted by L2-norm from the cost function gradient at the current iteration with

$$\mathbf{H}_k^{-1} = \frac{\left\|\nabla E\left(\mathbf{m}_k\right)\right\|^2}{\left\|\mathbf{H}_k^{-1}\nabla E\left(\mathbf{m}_k\right)\right\|^2}\mathbf{H}_k^{-1}. \tag{19}$$

Equation (19) is inserted into (9), resulting in the PCG.

2.4. The Workflow of the PCG-Based FWI.

Combining the principles of the core part of the Love wave FWI described in Sections 2.1–2.3, we give the detailed workflow of the PCG-based Love wave FWI as follows (Figure 1):

(1) An initial velocity model, the trial parameters (e.g., record time, maximum iteration number, and empirical parameter ε), and the iteration termination criteria are set.

(2) Information about the model velocity is input, equation (1) is solved to obtain a forward-propagating wavefield, and predicted seismic datasets are acquired.

(3) The predicted seismic datasets are matched with the observed data, the cost function, and the wavefield residual calculated with equation (3).

(4) A wavefield residual is set as the seismic source, a backward-propagating residual wavefield is initiated, and a backward-propagating wavefield is acquired.

(5) Advantage is taken of the calculated forward and backward propagating wavefields to determine the gradient with equations (13) and (14), and then the pseudo-HM is obtained with equations (16) and (17).

(6) Steps (2)–(5) are repeated until the gradient, and the pseudo-HM are calculated and stacked from all shots.

(7) A preconditioned operator is calculated with the calculated stacked pseudo-HM using equations (18) and (19).

(8) The calculated stacked gradient and the preconditioned operator are used to determine the scalar parameters with equations (9) and (10).

(9) The optimal step size is then calculated with the parabolic fitting method.

(10) The velocity model is updated by equation (9) with a calculated stacked gradient, preconditioned operator, and optimal step size.

(11) It is judged whether the iteration convergence condition or the maximum iteration number is met. The velocity model is output if either criterion is met; otherwise, Steps (2)–(10) are repeated.

3. Model Tests

The PCG and the CG are simultaneously applied to the noise-free and noise-containing tests of the two models to verify the antinoise capability and the aforementioned advantages of the PCG over the CG. Because the Love wave is much more sensitive to the S-wave velocity than to the density [15], only the S-wave velocity is updated in the inversion, and the density is not updated. In all tests, the density of the model is 1800 kg/m³, and the density and the source wavelet are treated as known quantities. The coordinates of the first shot and the first receiver are (0 m, 0 m). In total, 2000 points are sampled in the time domain with a sampling interval of 0.0002 s, resulting in a total seismic record of 0.4 s. The maximum number of iterations is set to 40. The iteration is terminated automatically once the decline of the cost function value of the two adjacent iterations is less than 0.00001 times the initial value ($k = 0$). To quantitatively assess the inversion result precision and to compare the strengths and weaknesses among different methods, a root-mean-square error (RMSE) with the following form is introduced:

$$\mathrm{RMSE} = \frac{\left\|\mathbf{m}_{\mathrm{inv}} - \mathbf{m}_{\mathrm{true}}\right\|_2}{\left\|\mathbf{m}_{\mathrm{true}}\right\|_2}, \tag{20}$$

where $\mathbf{m}_{\mathrm{inv}}$ and $\mathbf{m}_{\mathrm{true}}$ are the inversion results and the true values of the S-wave velocity model, respectively.
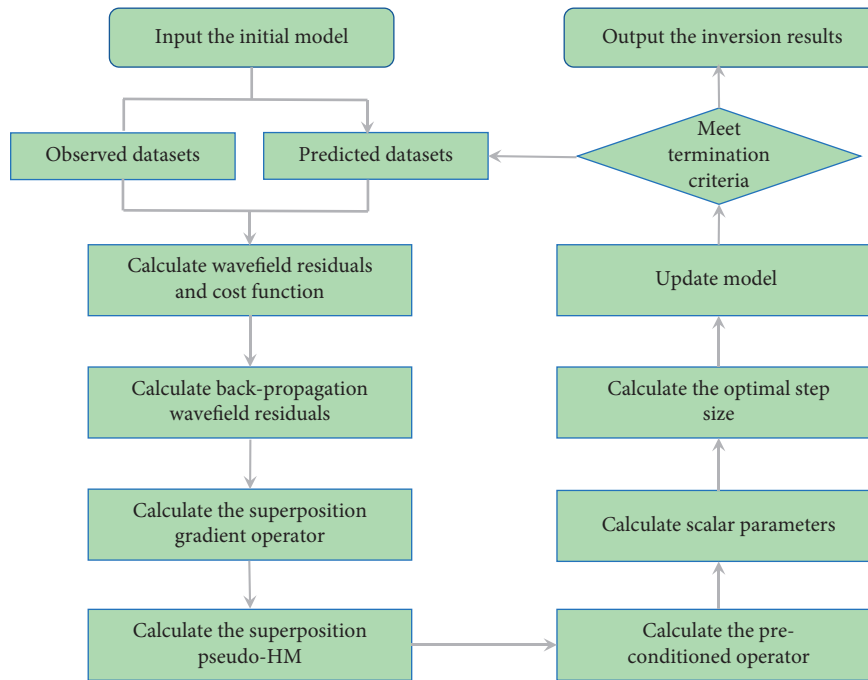
FIGURE 1: The workflow of the PCG-based FWI implementation.

3.1. Noise-Free Data. The first test example is a checkerboard model (Figure 2). The grid has a size of 41×101 in the vertical and horizontal directions with a step size of 0.5 in either direction. Hence, the actual model size is 30×40 m. The checkerboard is composed of 5×10 m blocks, and the S-wave velocity of each block is 300 m/s or 500 m/s, determined by the different positions (Figure 2(a)). The S-wave velocities of 300 m/s to 500 m/s are very common on a shallow subsurface. All 21 shots and 101 horizontal-component receivers (the white arrows and the solid yellow circles in Figure 2(a) represent the sources and receivers, respectively) are deployed on the surface in the *x*-direction with intervals of 2.5 m and 0.5 m. The initial model is set to a half-space with the S-wave velocity of 400 m/s (Figure 2(b)). The checkerboard model causes the transmitted seismic waves to scatter violently. When the seismic wave propagates to the boundary of each checkerboard block, the apparent scattered wave is generated due to the sudden change of the wave impedance, which can be clearly observed (Figure 3(a)). Furthermore, the first break Love wave with a shorter travel time than the scattered waves can also be observed (Figure 3(a)). These two sets of waves occupy most of the wavefield's energy ("L" and "S" in Figure 3(a) indicate these two seismic events). Scattered waves are ubiquitous in shallow subsurface exploration, such as the existence of cavities or high-velocity anomalies in the ground.

Because of the insufficient illumination and uneven energy distribution of the gradient operators, the normalized descent direction in the first iteration of the CG presents artifacts near the free surface of the model, and its amplitude decreases sharply in the deeper medium (Figure 2(c)), which seriously affects the updating of the deep medium parameters and the stability of the inversion process. Because the

PCG calculates the pseudo-HM of the cost function, the newly calculated descent direction contains the curvature information and balances the energy of the gradient operator so that it has the uniform amplitudes of the orders of magnitude in the shallow and deep parts of the model (Figure 2(d)). This improves the inversion accuracy of the deep medium, which suggests that the PCG can strengthen the wavefield's illumination capability and smooth the artifacts caused by double scattering and source coupling. Figures 2(e) and 2(f) present the inversion results of the CG and the PCG. The CG can accurately reconstruct shallow checkerboard blocks and roughly display their boundaries, yet it cannot reconstruct deep checkerboard blocks (Figure 2(e)). Figure 3(g) also shows that the waveforms of inversion results fit poorly with the observed waveforms (the part with the worst fitting degree is enclosed by a green square and displayed in a magnified display). In contrast, the PCG reconstructs the model more precisely (Figure 2(f)), which is intuitively discernable in the S-wave velocity profiles in Figures 3(c) and 3(d). It is noted that due to the issue of the overestimation and underestimation of model parameters, the final inversion results have an upper bound of the velocity in the deep layer that is greater than 500 m/s. In comparison, the lower bound of the velocity at the surface is less than 300 m/s (Figures 3(c) and 3(d)). This is the phenomefnon of illumination excess, which is caused by overadjusting the gradient operator by the preconditioned operator during the iteration. Therefore, a constraint criterion for the model updates should be added in the iterative process to solve this issue. In the optimization process, the PCG has a faster convergence rate than the CG, and the cost function value and the RMSE are lower throughout (Figures 3(e) and 3(f)). The final value of the
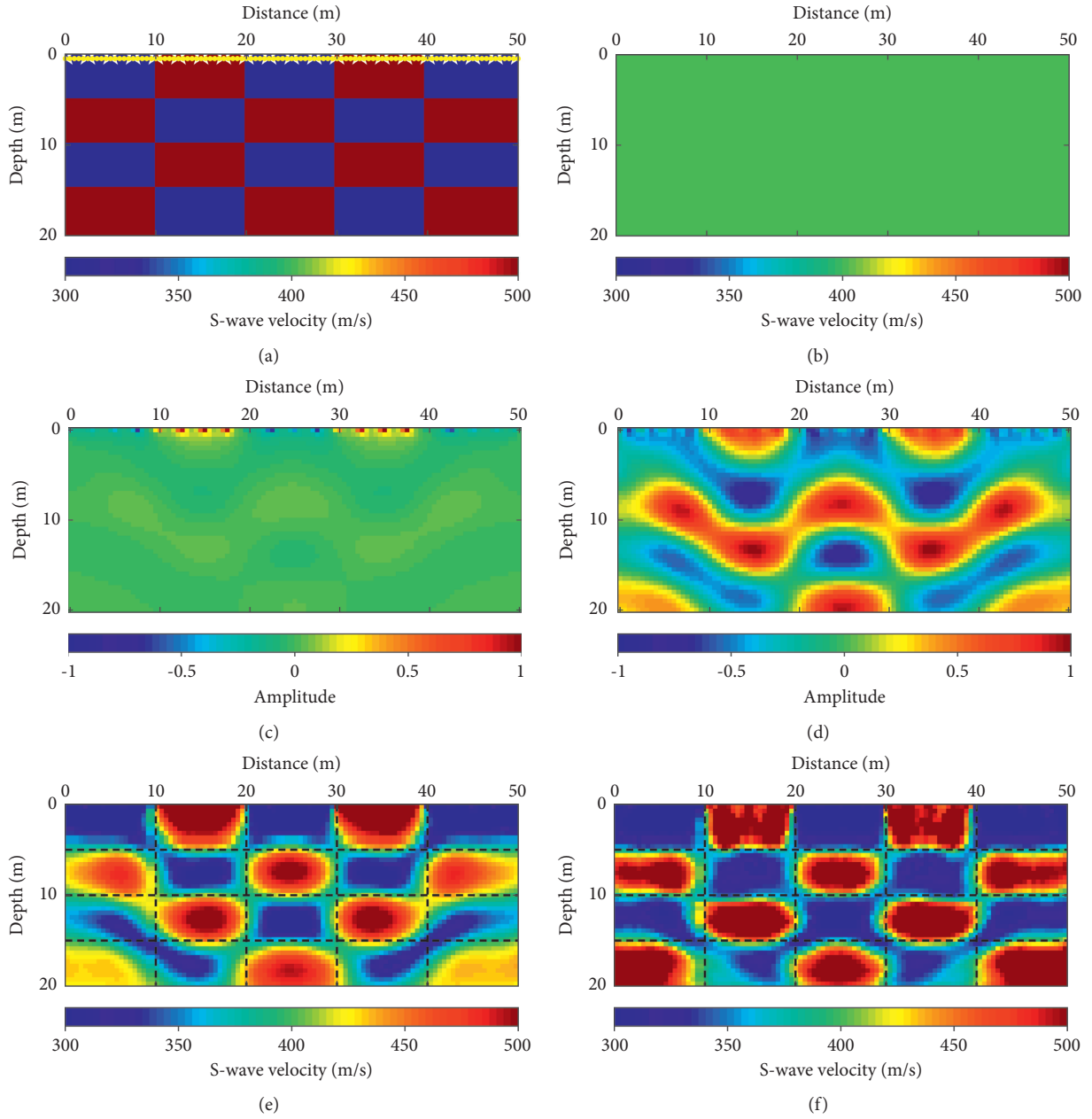
FIGURE 2: Noise-free data test for the checkerboard model. (a) The true model. The white arrows and the solid yellow circles represent the positions of the shots and the receivers. (b) The initial model. (c-d) The normalized descent directions of the CG and the PCG in the first iteration. (e-f) The inversion results of CG and the PCG. The black dotted lines represent the boundary where the S-wave velocity changes in the true model.

normalized cost function value (misfit) of the two algorithms is close to 0 (Table 1), and the waveforms of the observed data and the inversion results are nearly identical (Figure 3(h)). However, there are still some differences between the reconstructed model and the true model, and the final RMSE values of the PCG and the CG are 0.1169 and 0.16, respectively (Table 1). It is suspected that the reason for this phenomenon is either that the initial model is too poor and it is not in the neighborhood of the optimal global solution, there is cycle-skipping, or there are numerical errors.

In order to further illustrate the advantages of the PCG in enhancing the illumination capability and improving the imaging accuracy of deep media, the second test example is a deeper model containing a complex high-velocity anomaly and a fault (Figure 4). The grid in this model has a size of 81×101 in the vertical and horizontal directions with a step length of 0.5 m in either direction, resulting in an actual model size of 40×50 m. The model has three layers: the first layer has a velocity of 300 m/s, the second layer has a velocity of 400 m/s and a "boot-like" high-velocity anomaly, and the third layer has a velocity of 500 m/s and a fault (Figure 4(a)).
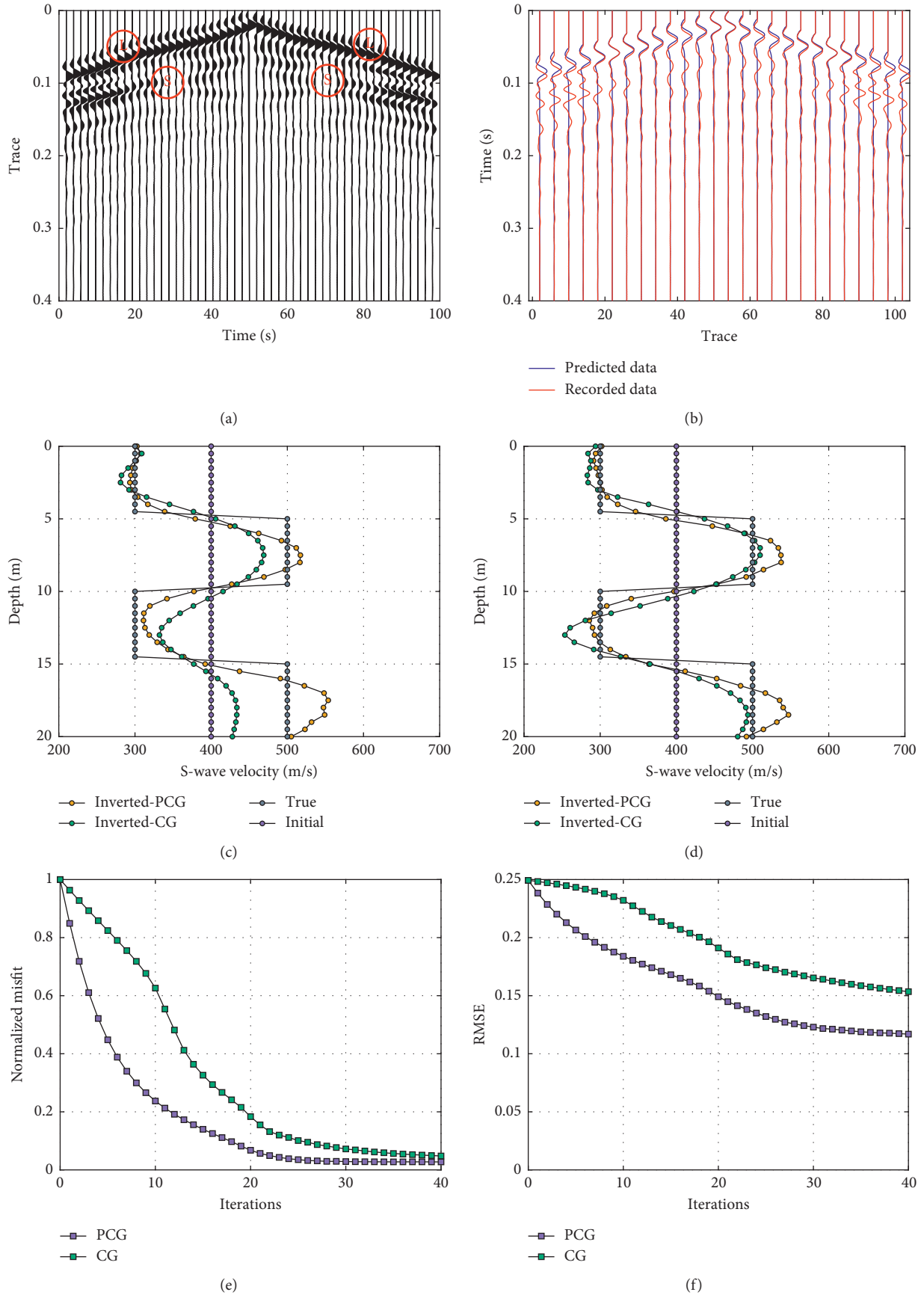
(a)

(b)

(c)
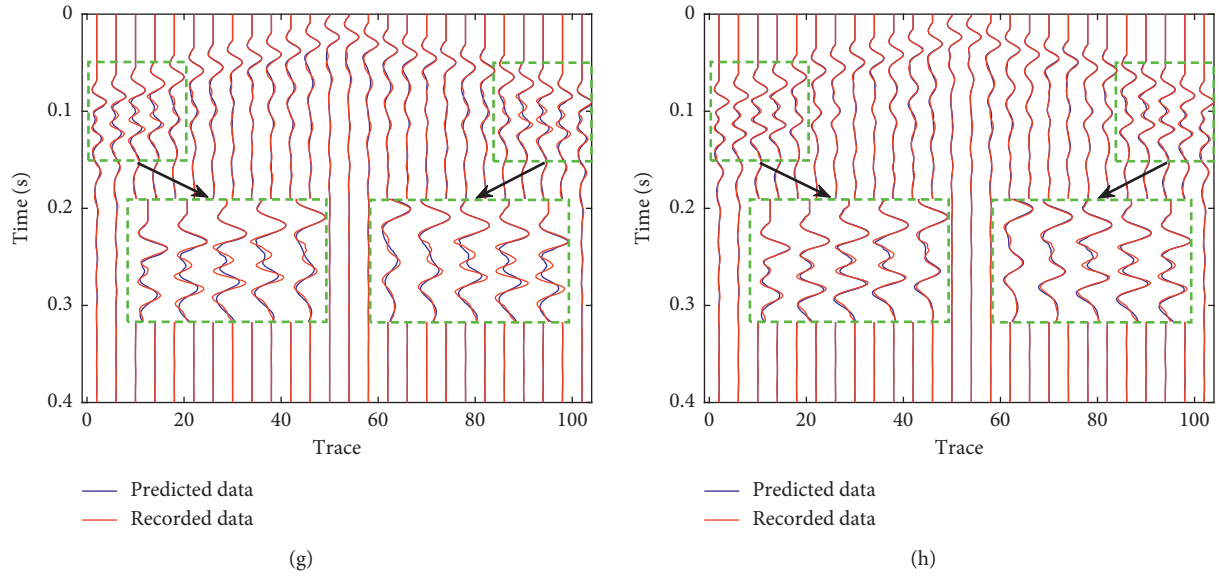
(d)

(e)

(f)

FIGURE 3: Continued.

(g)

(h)

FIGURE 3: Normalized seismograms and evaluation curves for the checkerboard model test. (a) The normalized seismograms of shot 11 from the true model. "L" and "S" indicate the first break Love wave and scattered wave events, respectively. (b) Comparison between the observed (red) and forward (blue) waveforms of shot 1 for the true and initial model. (c-d) The S-wave velocity vertical profiles of the true model, initial model, and inversion results of the PCG and CG at distance = 5 m and distance = 25 m, respectively. (e) Normalized misfits of PCG and CG versus iteration number. (f) The RMSE for the inversion results and the true value of the PCG and CG versus the iteration number. (g) Comparison between the observed (red) and forward (blue) waveforms of shot 1 of the CG inversion results. (h) Comparison between the observed (red) and forward (blue) waveforms of shot 1 of the PCG inversion results.

TABLE 1: Comparative evaluation of the CG and PCG performances in the checkerboard model noise-free test.

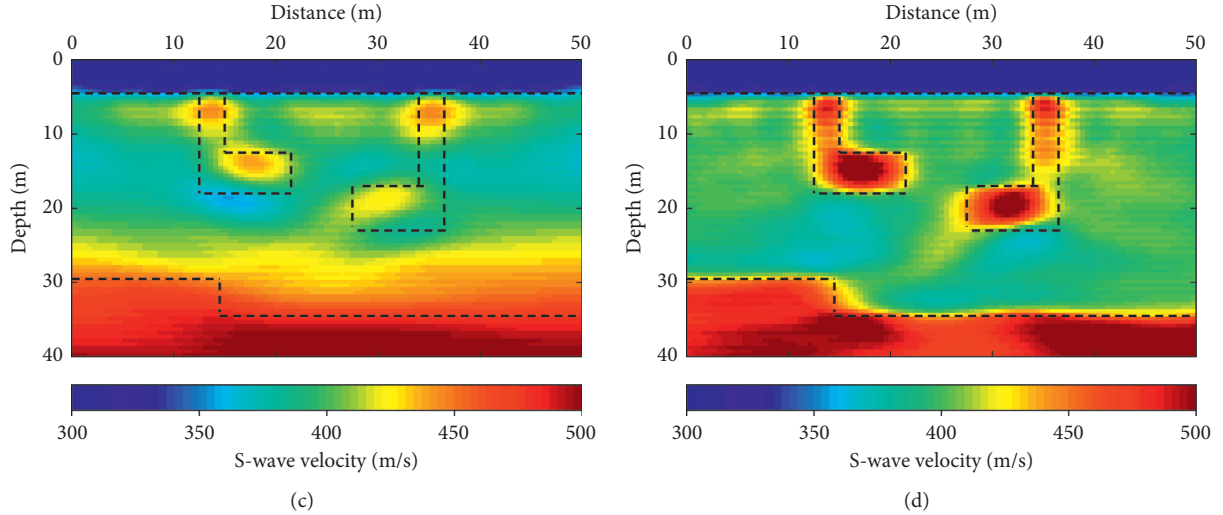| Optimization algorithm	Total iterations	Normalized final misfit value	Initial value of RMSE	Final value of RMSE
CG	40	0.0480	0.2493	0.1535
PCG	40	0.0275	0.2493	0.1169

(a)

(b)
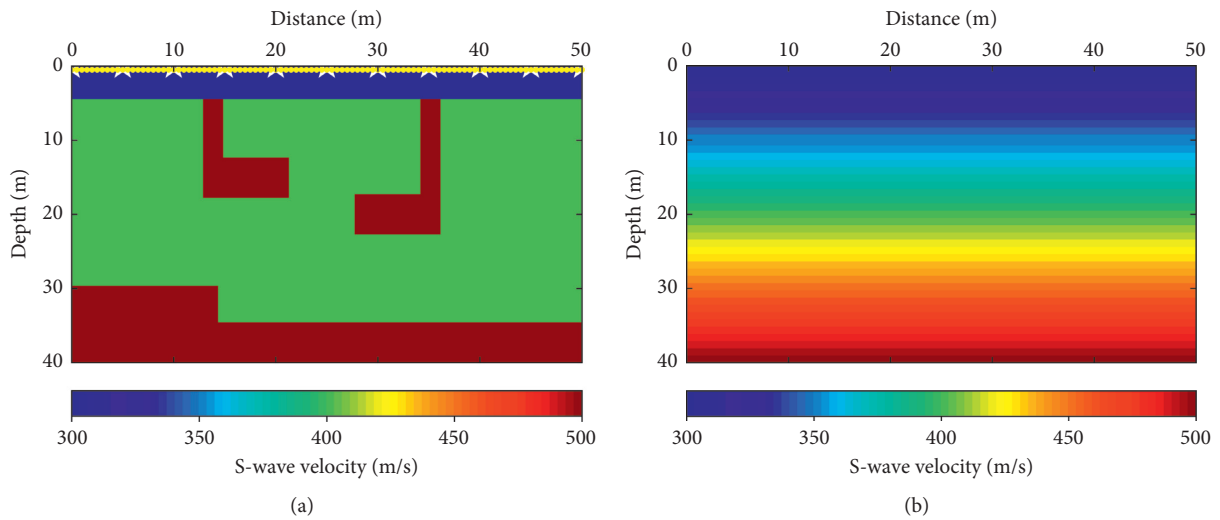
FIGURE 4: Continued.

FIGURE 4: Noise-free data test of the complex structure model. (a) The true model. The white arrows and the solid yellow circles represent the positions of the shots and the receivers. (b) The initial model. (c-d) The inversion results of the CG and the PCG. The black dotted lines represent the boundary where the S-wave velocity changes in the true model.

All 11 shots and 101 horizontal-component receivers (the white arrows and solid yellow circles in Figure 4(a) represent sources and receivers, respectively) are deployed on the surface in the x-direction with intervals of 5 m and 0.5 m, respectively. The initial model without information about the anomalies is set as a gradient model with 40 layers, with a thickness of 1 m per layer and a velocity of 300 m/s for the first layer, and the velocity gradient is 5 m/s (Figure 4(b)). To eliminate the large amplitude of the gradient due to the seismic source coupling effect at a free surface in the CG, the gradient is preprocessed by setting the gradient to zero at the grids that are at the free surface and one line below [35], making our demonstration more convincing. Some of the reflected waves produced by changes in the wave impedance can be observed ("R" in Figure 5(a) indicates the seismic event).

The CG inversion results precisely reconstruct the shallow morphology and display part of the high-velocity anomalies, yet they cannot be used to rebuild the fault structure (Figure 4(c)). The waveforms of the inversion results are basically fitted to the observed waveforms. However, the waveforms of the reflected wave are poorly matched (Figure 5(g), the poorly matched parts have been circled and enlarged in green squares). In contrast, the PCG inversion results reconstruct the anomalies and structures more precisely and clearly delineate the boundaries (Figure 4(d)), and the final RMSE value reaches 0.0320 (Figure 5(f) and Table 2). Both parameter overestimation and underestimation still exist (Figures 5(c) and 5(d)). Because the updating of the shallow part of the model contributes greatly to the function value, the convergence rate of the CG is similar to that of the PCG in the first ten iterations (gradient zeroing near the free surface plays a role). Later, due to the insufficient updating of the deep medium parameters, the CG stops the optimization. Because the PCG makes it possible for the deep medium parameters to be updated due to better illumination capability at depth, the cost function value continues to

decline. The final value is 0.0064, which is less than 1/9 of that in the CG (Figure 5(e) and Table 2). The waveforms of the inversion results match well with the observed waveforms (Figure 5(h)). This comparison also confirms the importance of the HM for improving the inversion precision and proves that the PCG is exceptionally effective. In this research, we suggest using the preconditioned gradient-based algorithm such as the PCG in FWI.

3.2. Noise-Containing Data. We apply the PCG to the second test example described above, but 20% and 50% Gaussian white noise is added to the observed data (Figures 6(a) and 6(b)) to verify the strong antinoise capability of the preconditioned gradient-based algorithm. The initial model is the same as that in the noise-free test (Figure 4(b)). The inversion results are shown in Figures 6(c) and 6(d). Even if the observed data contain noise, good inversion results can still be obtained, with which the high-velocity anomaly and the fault can be reconstructed (Figures 6(c) and 6(d)). Because the noise-containing waveforms cannot match the noise-free waveforms, in contrast to the value of the normalized cost function converging close to almost 0 in the noise-free test, the normalized cost function values of the noise test (Figure 7(a)) converge to 0.6725 and 0.8784 when the noise levels are 20% and 50%, respectively (Table 3). The inversion accuracy decreases with the increment of the noise level. However, when 50% noise is included in the observed data, the PCG final RMSE value is still lower and the inversion accuracy is even higher than that of the noise-free CG inversion (Figure 7(b) and Table 3). Additionally, the higher the noise level is, the more serious the parameter overestimation is (Figures 7(c) and 7(d)), which may be caused by the boundary effect. The noise-containing data test shows that the preconditioned gradient-based algorithm has high stability even if the observed data are contaminated by noise.
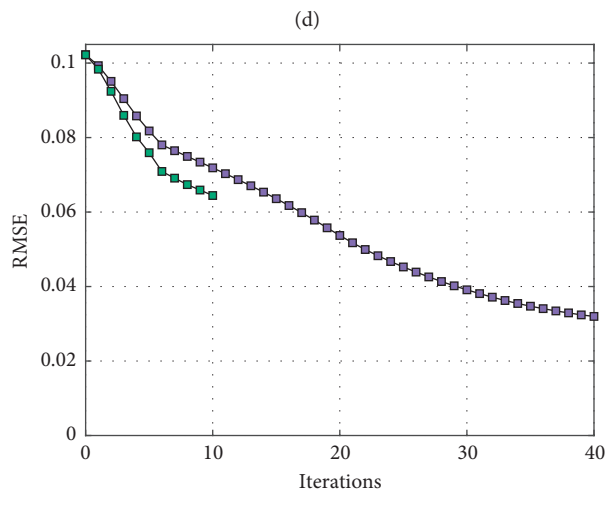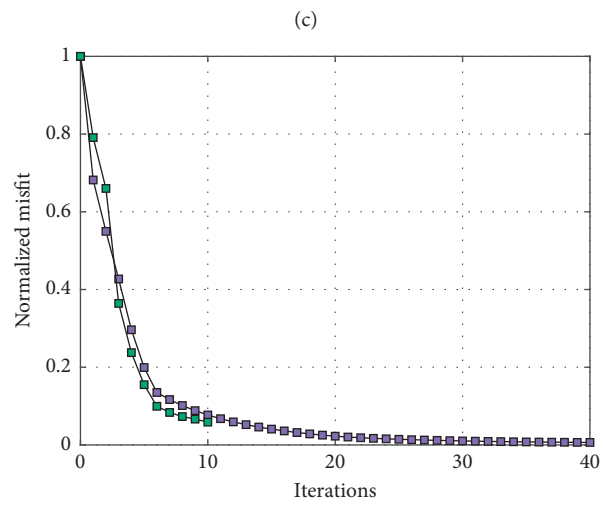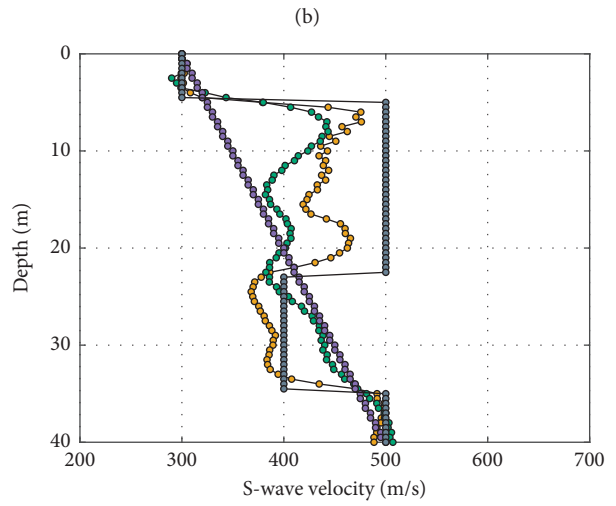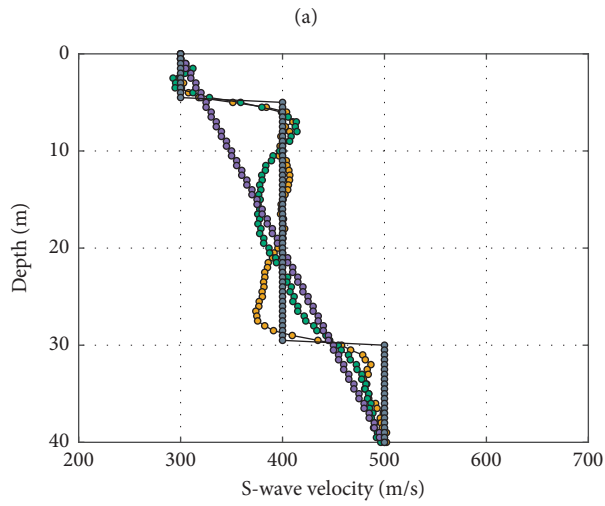
(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)
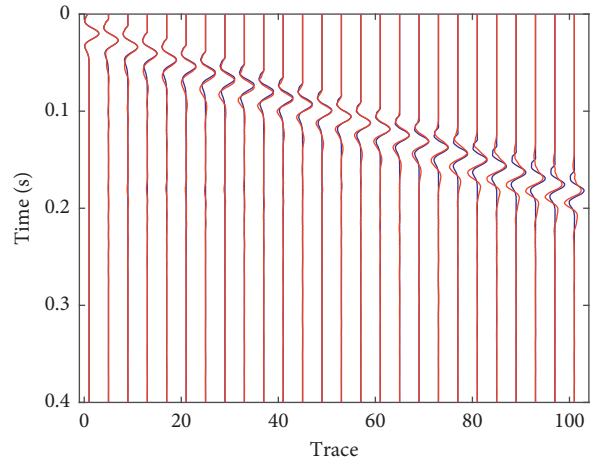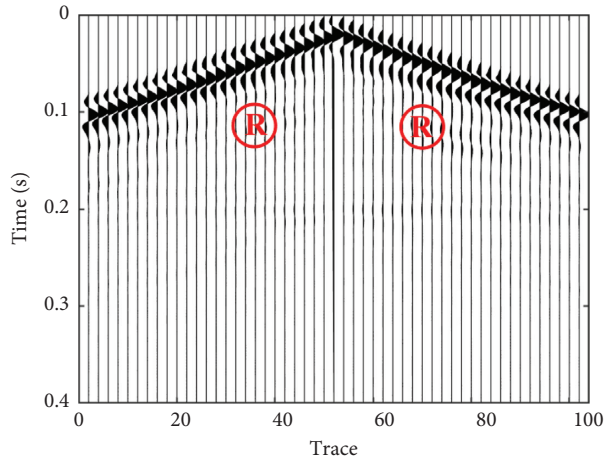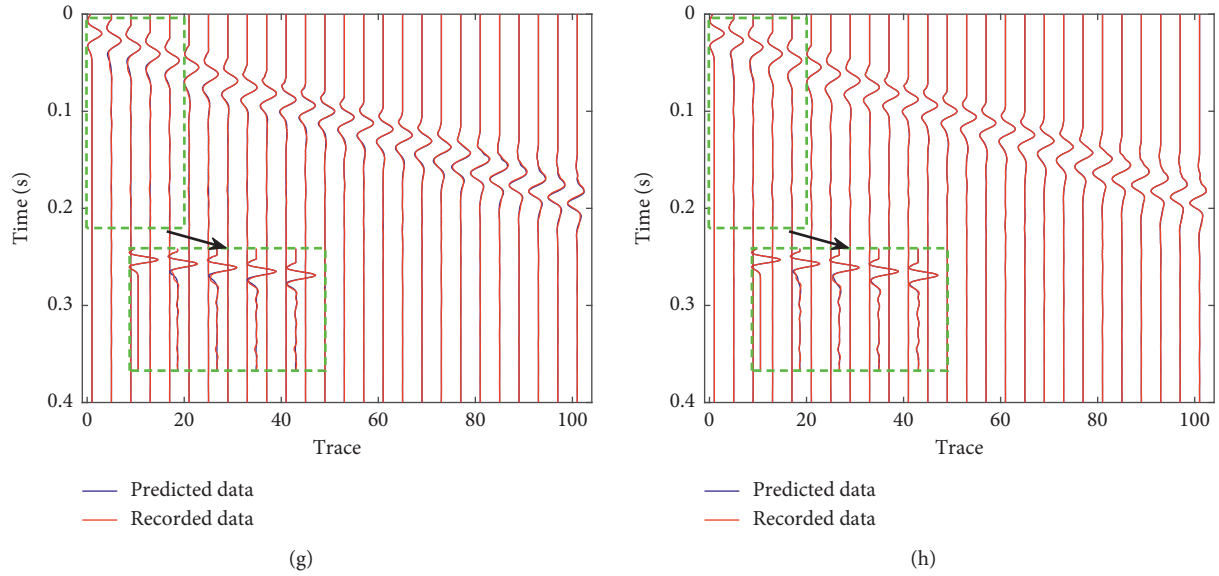
FIGURE 5: Normalized seismograms and evaluation curves for the complex structure model test. (a) The normalized seismograms of shot 6 from the true model. "R" indicates the reflected Love wave. (b) Comparison between the observed (red) and forward (blue) waveforms of the true and initial model. (c-d) The 1D S-wave velocity vertical profiles of the true model, initial model, and inversion results of the PCG and CG at distance = 10 m and distance = 35 m, respectively. (e) Normalized misfits of the PCG and CG versus iteration number. (f) The RMSE between the inversion results and the true values of the PCG and CG versus the iteration number. (g) Comparison between the observed (red) and forward (blue) waveforms of shot 1 of the CG inversion results. (h) Comparison between the observed (red) and forward (blue) waveforms of shot 1 of the PCG inversion results.

TABLE 2: Comparative evaluation of the CG and PCG performances in the complex structure model noise-free test.

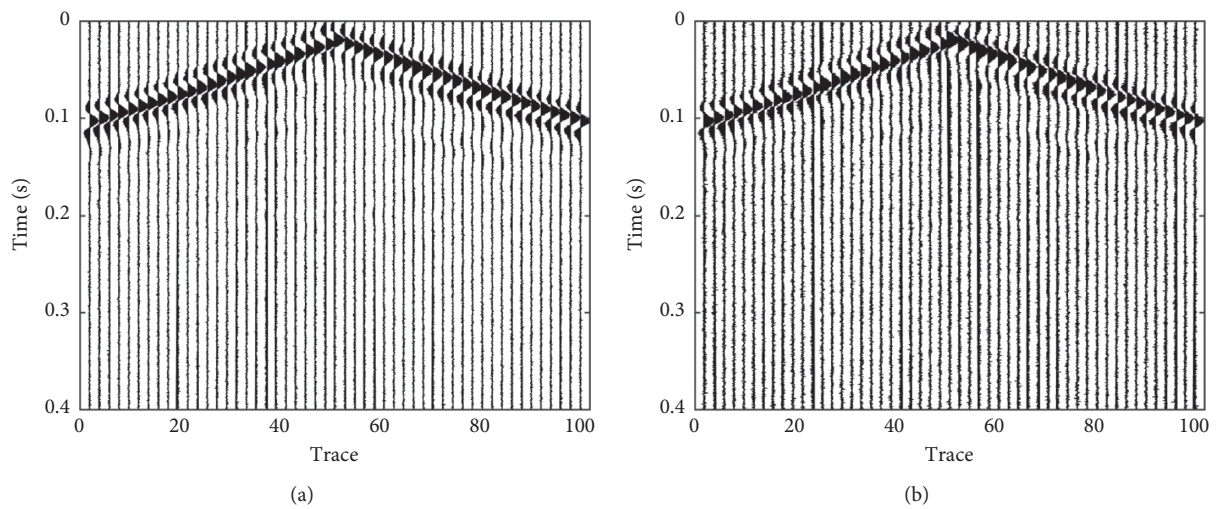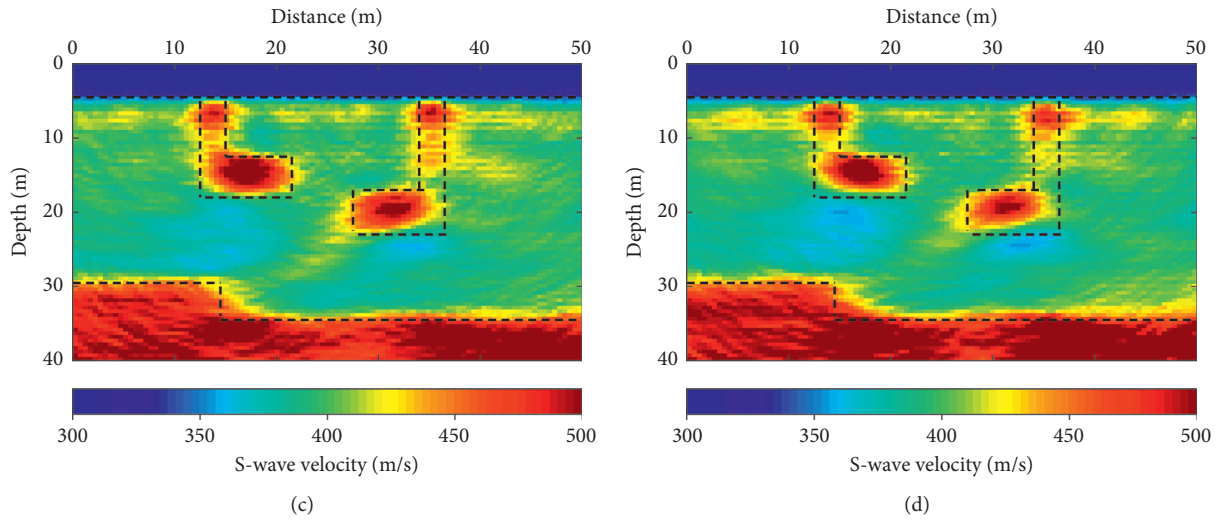| Optimization algorithm	Total iterations	Normalized final misfit value	Initial value of RMSE	Final value of RMSE
CG	10	0.0587	0.1022	0.0644
PCG	40	0.0064	0.1022	0.0320

(a)

(b)

FIGURE 6: Continued.

(c)

(d)

FIGURE 6: Normalized seismograms and inversion results based on the PCG in the noise-containing test of the complex structure model. (a-b) Normalized seismograms with 20% and 50% Gaussian white noise for shot 6 from the true model. (c-d) The inversion results for the normalized seismograms with 20% and 50% Gaussian white noise.
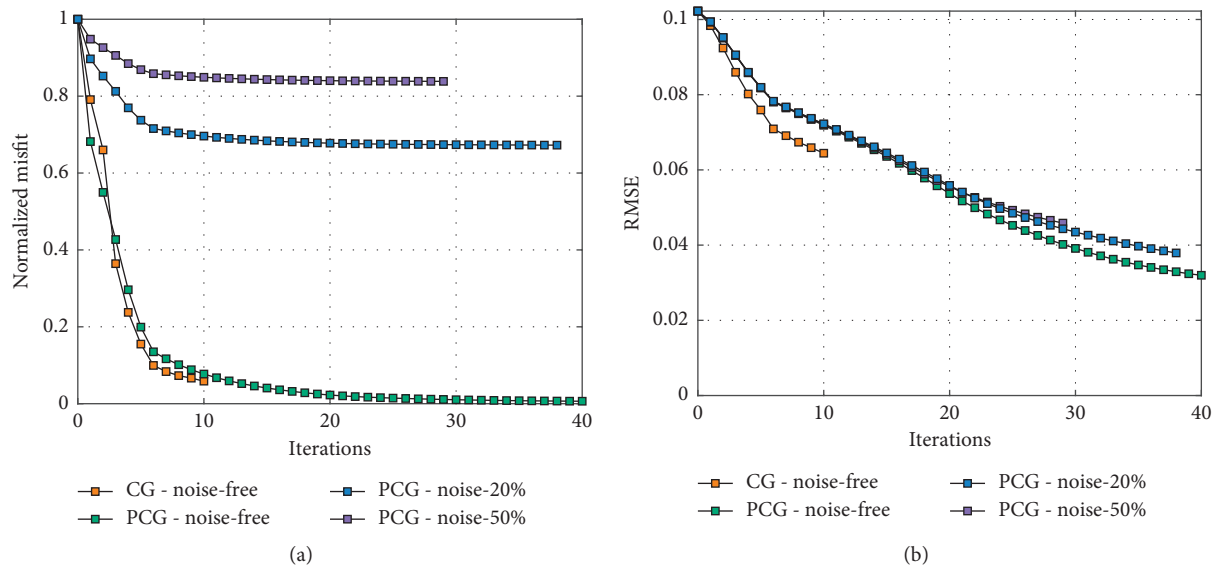

(a)

(b)

FIGURE 7: Continued.

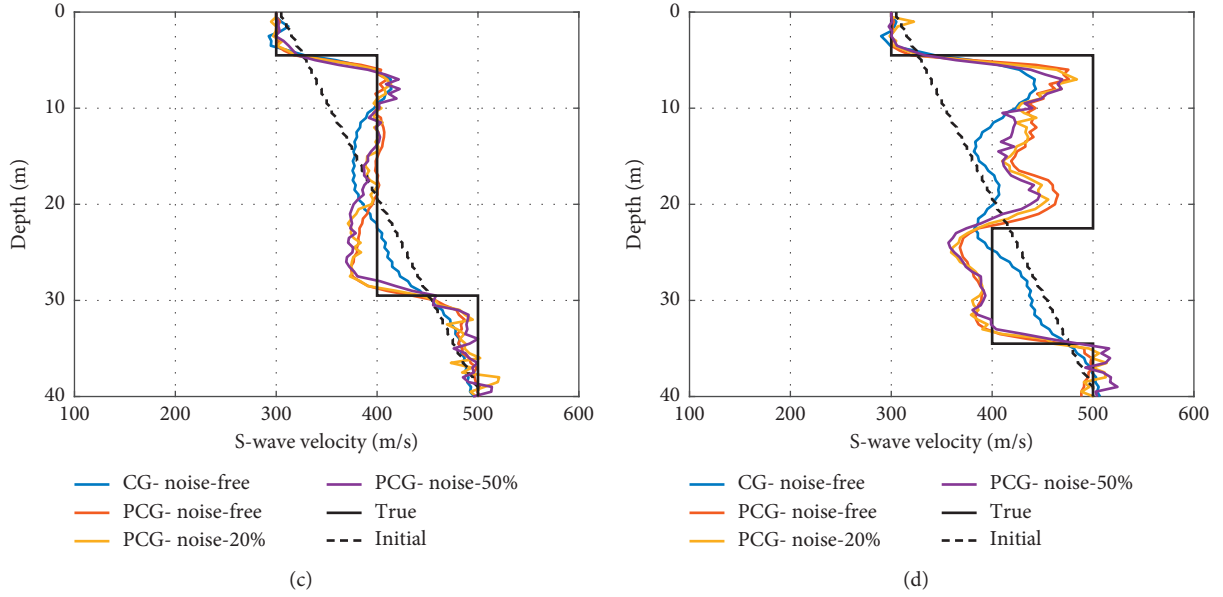(c)   (d)

FIGURE 7: Evaluation curves for the complex structure model noise-containing test. (a) Normalized misfits for the CG when the seismograms contain no noise and for the PCG when the seismograms contain no noise, 20% noise, and 50% noise versus the iteration number. (b) The RMSE for the CG when the seismograms contain no noise and for the PCG when the seismograms contain no noise, 20% noise, and 50% noise versus the iteration number. (c-d) The 1D S-wave velocity vertical profiles of the true model, initial model, and inversion results at distance $= 10$ m and distance $= 35$ m.

TABLE 3: Comparative evaluation of the PCG performance for seismograms with 20% and 50% Gaussian white noise in the complex structure model noise-containing test.

Noise level (%)	Total iterations	Normalized final misfit value	Initial value of RMSE	Final value of RMSE
20	38	0.6725	0.1022	0.0379
50	29	0.8784	0.1022	0.0459

4. Discussion and Conclusion

With reference to the theory of inverse scattering, we develop a new pseudo-HM-based preconditioned approach for the 2D FWI of a Love wave to make full use of the wavefield information contained in the HM but not contained in the gradient. In this novel method, the stress components of the Love wave wavefield are used to construct a pseudo-inverse HM to filter, balance, and regularize the gradient. This reduces the influence of the wavefield propagation effects, including the double scattering, transmission, and geometric diffusion on the gradient, in order to improve the inversion results. In addition, because this method is independent of each component of the backward-propagating wavefield, it will not affect the focus of the data residual for the model perturbation, and the method will ensure that the gradient is not distorted. The stress components required by our new method have been obtained in the necessary forward simulation during each iteration, so the method does not have extra calculation costs and it ensures the calculation efficiency. The synthesis tests show that our new method can significantly accelerate the convergence rate, suppress the artifacts near the surface, and improve the model reconstruction accuracy, especially in the deep region of the model that is more affected by the defocus effect and geometric spreading. Furthermore,

the PCG also has a strong antinoise ability that is very important for processing field datasets, and we believe that the pseudo-inverse HM plays a key role in this. However, the pseudo-inverse HM overcompensates for the gradient in some cases, resulting in overillumination of the wavefield and overestimation of the parameters, especially near the surface and in the deep region of the model. Therefore, the confining criterion of the model update must be added in the iterative optimization, which requires further study. In addition, only the ability of the PCG to resist Gaussian white noise is tested in the synthetic tests, but some special noises such as traffic and electromagnetic noise may exist in the actual situation. This still requires further testing and discussion. At the end, similar to the application of the PCG to the Love wave FWI described in this paper, a similar method can be applied to a Rayleigh wave and other body waves. It should be noted that the specific form of the proposed pseudo-HM is not invariable for different types of waves, and specific derivation and construction should be carried out according to different wave equations in combination with equation (15).

Data Availability

The data and the code in this paper can be obtained by contacting the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have influenced the work reported in this paper.

Acknowledgments

References

[1] C. B. Park, R. D. Miller, and J. Xia, "Multichannel analysis of surface waves," *Geophysics*, vol. 64, no. 3, pp. 800–808, 1999.

[2] J. Xia, Y. Xu, Y. Luo, R. D. Miller, R. Cakir, and C. Zeng, "Advantages of using multichannel analysis of love waves (MALW) to estimate near-surface shear-wave velocity," *Surveys in Geophysics*, vol. 33, no. 5, pp. 841–860, 2012.

[3] T. Bohlen, S. Kugler, G. Klein, and F. Theilen, "1.5D inversion of lateral variation of scholte-wave dispersion," *Geophysics*, vol. 69, no. 2, pp. 330–344, 2004.

[4] J. Li, Z. Feng, and G. Schuster, "Wave-equation dispersion inversion," *Geophysical Journal International*, vol. 208, no. 3, pp. 1567–1578, 2017.

[5] J. Li, S. Hanafy, Z. Liu, and G. T. Schuster, "Wave-equation dispersion inversion of love waves," *Geophysics*, vol. 84, no. 5, pp. R693–R705, 2019.

[6] A. Tarantola, "Inversion of seismic reflection data in the acoustic approximation," *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984.

[7] A. Tarantola, "Inverse problem theory. methods for data fitting and parameter estimations," *Geophysical Journal*, vol. 94, pp. 167-168, 1988.

[8] Y. Pan, J. Xia, Y. Xu, L. Gao, and Z. Xu, "Love-wave waveform inversion in time domain for shallow shear-wave velocity," *Geophysics*, vol. 81, no. 1, pp. 1JF–Z7, 2016.

[9] J. Woodon, T. Constantinos, and K. Y. Seo, "Full waveform inversion with angle-dependent gradient preconditioning using wavefield decomposition," *Journal of Applied Geophysics*, vol. 159, pp. 23–31, 2018.

[10] L. Liu, H. Liu, and H. Zhang, "Full waveform inversion based on modifien quasi-Newton equation," *Chinese Journal of Geophysics*, vol. 56, no. 7, pp. 2447–2451, 2013.

[11] R. G. Pratt and M. H. Worthington, "Inverse theory applied to multi-source cross-hole tomography.. part 1: acoustic wave-equation method1," *Geophysical Prospecting*, vol. 38, no. 3, pp. 287–310, 1990.

[12] L. Métivier, R. Brossier, J. Virieux, and S. Operto, "The truncated Newton method for full waveform inversion," *SEG, Expanded Abstracts*, Society of Exploration Geophysicists, Tulsa, Oklahoma, 2012.

[13] L. Métivier, R. Brossier, and J. Virieux, "Full waveform inversion and the truncated Newton method," *SIAM Journal on Scientific Computing*, vol. 35, no. 2, pp. B401–B437, 2013.

[14] L. Métivier, F. Bretaudeau, R. Brossier, J. Virieux, and S. Operto, "Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures," *Geophysical Prospecting*, vol. 62, pp. 1353–1375, 2014.

[15] J. Xia, *High-frequency Surface Wave Method (In Chinese)*, Press of China University of Geosciences, Wuhan, China, 2015.

[16] R. Eslick, G. Tsoflias, and D. Steeples, "Field investigation of love waves in near-surface seismology," *Geophysics*, vol. 73, no. 3, pp. G1–G6, 2008.

[17] P. Mora, "Nonlinear two-dimensional elastic inversion of multioffset seismic data," *Geophysics*, vol. 52, no. 9, pp. 1211–1228, 1987.

[18] R. E. Plessix, "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications," *Geophysical Journal International*, vol. 167, no. 2, pp. 495–503, 2006.

[19] O. Gauthier, J. Virieux, and A. Tarantola, "Two-dimensional nonlinear inversion of seismic waveforms: numerical results," *Geophysics*, vol. 51, no. 7, pp. 1387–1403, 1986.

[20] Y. Wang and Y. Rao, "Reflection seismic waveform tomography," *Journal of Geophysical Research*, vol. 114, no. B3, Article ID B03304, 2009.

[21] A. J. Brenders and R. G. Pratt, "Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model," *Geophysical Journal International*, vol. 168, no. 1, pp. 133–151, 2007.

[22] R. G. Gerhard Pratt, C. Shin, and G. J. Hicks, "Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion," *Geophysical Journal International*, vol. 133, no. 2, pp. 341–362, 1998.

[23] R. Chen and K. T. Tran, "2D Gauss-Newton full waveform inversion of SH- and love waves in the time domain," *Journal of Applied Geophysics*, vol. 191, Article ID 10436, 2021.

[24] K. T. Tran, M. Mirzanejad, M. McVay, and D. Horhota, "3-D time-domain Gauss-Newton full waveform inversion for near-surface site characterization," *Geophysical Journal International*, vol. 217, no. 1, pp. 206–218, 2019.

[25] C. Shin, S. Jang, and D.-J. Min, "Improved amplitude preservation for prestack depth migration by inverse scattering theory," *Geophysical Prospecting*, vol. 49, no. 5, pp. 592–606, 2001.

[26] Y. Choi, D.-J. Min, and C. Shin, "Frequency-domain elastic full waveform inversion using the new pseudo-Hessian matrix: experience of elastic marmousi-2 synthetic data," *Bulletin of the Seismological Society of America*, vol. 98, no. 5, pp. 2402–2415, 2008.

[27] R. Brossier, S. Operto, and J. Virieux, "Seismic imaging of complex on shore structures by 2D elastic frequency-domain full-waveform inversion," *Geophysics*, vol. 74, no. 2, pp. WCC63–WCC76, 2009.

[28] Y. Rao, Y. Wang, and D. Han, "Seismic waveform tomography with simplified restarting scheme," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 1, pp. 135–139, 2019.

[29] W. Pan, K. A. Innanen, G. F. Margrave, and D. Cao, "Efficient pseudo-Gauss-Newton full-waveform inversion in the τ-p domain," *Geophysics*, vol. 80, no. 5, pp. R225–R14, 2015.

[30] J. Virieux, "SH-wave propagation in heterogeneous media: velocity-stress finite-difference method," *Geophysics*, vol. 49, no. 11, pp. 1933–1942, 1984.

[31] W. Rodi and R. L. Mackie, "Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion," *Geophysics*, vol. 66, no. 1, pp. 174–187, 2001.

[32] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, Berlin/Heidelberg, Germany, 2006.

[33] J. C. Gilbert and J. Nocedal, "Global convergence properties of conjugate gradient methods for optimization," *SIAM Journal on Optimization*, vol. 2, no. 1, pp. 21–42, 1992.

[34] C. Castellanos, V. Etienne, G. Hu, S. Operto, R. Brossier, and J. Virieux, "Algorithmic and methodological developments towards full waveform inversion in 3D elastic media," *SEG Technical Program Expanded Abstracts*, Society of Exploration Geophysicists, Tulsa, Oklahoma, pp. 2793–2798, 2011.

[35] D. Köhn, D. De Nil, A. Kurzmann, A. Przebindowska, T. Bohlen, and T. Bohlen, "On the influence of model parametrization in elastic full waveform tomography," *Geophysical Journal International*, vol. 191, no. 1, pp. 325–345, 2012.