Petroleum Science 19 (2022) 1582–1591

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Reflection-based traveltime and waveform inversion with second-order optimization

Teng-Fei Wang [a, b], Jiu-Bing Cheng [a, b, *], Jian-Hua Geng [a, b]

[a] State Key Laboratory of Marine Geology, Tongji University, Shanghai, 200092, China
[b] School of Ocean and Earth Science, Tongji University, Shanghai, 200092, China

ARTICLE INFO

Article history:
Received 2 September 2021
Accepted 10 February 2022
Available online 16 February 2022

Edited by Jie Hao

Keywords:
Reflection waveform inversion
Reflection traveltime inversion
Gauss-Newton
Hessian

ABSTRACT

Reflection-based inversion that aims to reconstruct the low-to-intermediate wavenumbers of the subsurface model, can be a complementary to refraction-data-driven full-waveform inversion (FWI), especially for the deep target area where diving waves cannot be acquired at the surface. Nevertheless, as a typical nonlinear inverse problem, reflection waveform inversion may easily suffer from the cycle-skipping issue and have a slow convergence rate, if gradient-based first-order optimization methods are used. To improve the accuracy and convergence rate, we introduce the Hessian operator into reflection traveltime inversion (RTI) and reflection waveform inversion (RWI) in the framework of second-order optimization. A practical two-stage workflow is proposed to build the velocity model, in which Gauss-Newton RTI is first applied to mitigate the cycle-skipping problem and then Gauss-Newton RWI is employed to enhance the model resolution. To make the Gauss-Newton iterations more efficiently and robustly for large-scale applications, we introduce proper preconditioning for the Hessian matrix and design appropriate strategies to reduce the computational costs. The example of a real dataset from East China Sea demonstrates that the cascaded Hessian-based RTI and RWI have good potential to improve velocity model building and seismic imaging, especially for the deep targets.

1. Introduction

Complex heterogeneities naturally existing in the Earth's subsurface lead to the spatial variations of the elastic properties, such as velocity, density and so on. Seismic wavefield propagates in the subsurface and carries important elastic responses that can be used for seismic imaging. In the exploration seismology, seismic tomography and imaging methods are employed to estimate the subsurface elastic properties, with which the type of pore fluid or the distribution of stress field can be predicted for the purpose of hydrocarbon reservoir exploration and geotechnical engineering.

Exploration seismic imaging is practically applied based on a scale separation of velocity model, i.e., a smooth background and a rough perturbation (Claerbout, 1971). The smooth background velocity is first estimated by the migration velocity analysis (MVA) or ray-based traveltime tomography and then the rough perturbation is obtained by the migration process (Claerbout, 1985). Over the past decades, one of the most important tasks for seismic imaging is to reconstruct a background velocity model that can correctly honor the kinematics of the wave propagation. The standard velocity model building for prestack-depth-migration (PSDM) is based on the reflection traveltime tomography with ray-tracing, in which the migration velocity model is iteratively updated by back-projecting the residual depth or moveout (RMO) on the common image gathers (CIGs) along the ray path (Stork, 1992; Wang and Pratt, 1997; Woodward, 2008). Due to the limitation of ray-based theory, the PSDM-based reflection traveltime tomography easily fails in the presence of complex velocity variation and lead to inaccurate imaging of the target. Besides, the generation of CIGs during the PSDM and the picking of RMO are quite expensive and clumsy to support multifold iterations of velocity model building (Alder et al., 2008; Guillum et al., 2008).

To overcome the limitation of ray-based method, wave-equation-based migration and tomography techniques, such as reverse-time migration (RTM) and full-waveform inversion (FWI), have been proved quite effective to improve the accuracy of seismic

* Corresponding author.
E-mail address: cjb1206@tongji.edu.cn (J.-B. Cheng).

https://doi.org/10.1016/j.petsci.2022.02.003

imaging (Virieux and Operto, 2009; Zhang et al., 2014). Nevertheless, FWI easily suffers from the cycle-skipping problems when starting model is poor and/or low frequency data is unavailable. Therefore, as a model building process, FWI only provides robust velocity update for the shallow part of the model where turning waves are sufficiently available. The velocity update in the deep part remains to be improved, especially in the field data applications. Xu et al. (2012) introduced wave-equation reflection waveform inversion (RWI) which aims to recover the background velocity by minimizing the waveform misfit between synthetic and observed reflections. In the context of wave-equation reflection tomography or waveform inversion, sharp contrasts of the elastic properties or reflectivity in the subsurface medium are prerequisite (provided by seismic migration), with which the reflection wavefields are generated to produce the tomographic misfit kernels ("rabbit-ear") connecting the shots and receivers to the reflectors. In the past few years, RWI becomes a quite promising method to recover the low-to-intermediate wavenumber components of the velocity model (Zhou et al., 2015; Wu and Alkhalifah, 2015; Wang and Cheng, 2017; Yao and Wu, 2017; Xu et al., 2019; Yao et al., 2020).

However, RWI still faces many challenges in practice. As a waveform fitting process, it is very likely to fall into the local minima when the initial model is far from the true one. A common strategy is to reformulate the inverse problem using traveltime-based objective function (Ma and Hale, 2013; Wang et al., 2015). Many recent works have shown that wave equation reflection traveltime inversion (RTI) can robustly recover the background velocity and thus allow the RWI starting from a better initial model (Brossier et al., 2015; Wang et al., 2018). The information content of the traveltimes and amplitudes are complementary, being sensitive to different features of the model (Wang and Pratt, 1997). Therefore, it is necessary to unite RTI and RWI to improve the reconstruction of the low-to-intermediate wavenumber features of the subsurface velocity model, e.g., Xu et al. (2019). Besides, all the aforementioned RTI and RWI approaches possess the issue of slow convergence because they only use the gradient-based first-order optimization methods. The second-order derivative of the objective function, i.e. the Hessian, implies the information of acquisition geometry, parameter trade-offs and the resolution ability of the observed data (Pratt et al., 1998; Operto et al., 2013; Pan et al., 2016; Wang and Cheng, 2017); therefore, it is a key point to improve the convergence of RWI. In the context of FWI, people use the diagonal or band diagonal Hessian to compensate the spherical divergence/geometric spreading and mitigate parameter trade-offs (Innanen, 2014; Shin et al., 2001; Wang et al., 2016). With full or approximate Hessian matrix, the quasi-Newton, Gauss-Newton or Newton method leads to better convergence (Brossier et al., 2009 Metivier et al. 2014, 2017; Liu et al., 2015; Pan et al., 2017). However, the studies of Hessian-based reflection tomography or waveform inversion are very insufficient. Recently, Wang et al. (2020) introduced the Hessian-based Gauss-Newton method to solve the RWI problem in acoustic media. Their attempt shows promising results for the reconstruction of the deep part velocity model.

In this paper, we will propose a practical workflow by cascading Hessian-based RTI and RWI, and provide some computational strategies to support large-scale applications. This paper is organized as follows: First, we will briefly review the Fréchet derivatives, functional gradient and Hessian matrix for reflection-based inversion based on the Born approximation. Then, we will introduce some practical strategies to precondition the Hessian-based inversion and deal with the computational challenges. Afterwards, we will demonstrate the two-stage Gauss-Newton RTI/RWI workflow with a real dataset from East China Sea. Finally, we draw some conclusions.

2. Reflection-based wave equation inversion

In the subsurface medium, the traveltime and waveform of seismic waves relate differently to the velocity perturbations. To better understand the reflection-based inverse problems, we first describe the forward problem in the framework of first-order Born approximation, and then review the Fréchet derivatives of reflection waveform and reflection traveltime. Accordingly, we derive the corresponding functional gradient and approximate Hessian matrices for RWI and RTI, which will be involved in the second-order optimization methods.

2.1. The reflection waveform sensitivity kernel

The constant-density acoustic wave equation in frequency domain can be written as:

$$A(\mathbf{m},\omega)U(\omega)=f(\omega), \tag{1}$$

where $A=-\frac{\omega^2}{v^2(\boldsymbol{x})}-\nabla^2$, ω is the frequency, v is the velocity, f is the source wavelet and U is the acoustic wavefield. For simplicity, the wavefield for a single source receiver pair can be expressed as:

$$U_0(\boldsymbol{r}|\boldsymbol{s})=G_0(\boldsymbol{r}|\boldsymbol{s})f(\boldsymbol{s}), \tag{2}$$

where r and s indicate the location of source and receiver, respectively, and $G_0=A^{-1}$ is the impulsive response for the background medium, i.e. Green's function. In seismic imaging, the velocity model can be split into a smooth background model and a rough perturbation model:

$$v=v_0+\delta v. \tag{3}$$

Then, for a given velocity perturbation, the Born approximation establishes a linear relation to the perturbation of wavefield:

$$\Delta U(\boldsymbol{r}|\boldsymbol{s})=\int 2k_0^2(\boldsymbol{x})\frac{\delta v(\boldsymbol{x})}{v_0(\boldsymbol{x})}G_0(\boldsymbol{r}|\boldsymbol{x})U_0(\boldsymbol{x}|\boldsymbol{s})\mathrm{d}\boldsymbol{x}, \tag{4}$$

with the background wavenumber $k_0=\frac{\omega}{v_0(\boldsymbol{x})}$. The above equation indicates that the wavefield perturbation at a certain receiver is the superposition of the wavefields emitting from the scatterers where the secondary source is scaled by the background wavefield. Particularly, if the scatterers in the subsurface are continuously distributed, they can be regarded as the rough perturbations or reflectivity. Then, the reflection wavefield satisfies:

$$\Psi(\boldsymbol{r}|\boldsymbol{s})=\int\varphi(\boldsymbol{r}|\boldsymbol{s})f(\boldsymbol{s})\mathrm{d}\boldsymbol{x}, \tag{5}$$

where $\varphi(\boldsymbol{r}|\boldsymbol{s})$ is the Born "wavepath" (Woodward, 1992):

$$\varphi(\boldsymbol{r}|\boldsymbol{s})=2k_0^2(\boldsymbol{x})G_0(\boldsymbol{r}|\boldsymbol{x})R(\boldsymbol{x})G_0(\boldsymbol{x}|\boldsymbol{s}) \tag{6}$$

and $R(\boldsymbol{x})$ is the reflectivity:

$$R(\boldsymbol{x})=\frac{\delta v(\boldsymbol{x})}{v(\boldsymbol{x})}. \tag{7}$$

For the RTI and RWI problems, we hope to create the linear relation between the background velocity and reflection wavefield perturbations. Taking the partial derivative of Ψ with respect to v_0 and making some rearrangements, we have the Fréchet derivative of reflection data:

$$\frac{\partial \Psi(\boldsymbol{r}|\boldsymbol{s})}{\partial v_0(\boldsymbol{x}')} = \int \frac{2k_0^2(\boldsymbol{x}')}{v_0(\boldsymbol{x}')} (\varphi(\boldsymbol{r}|\boldsymbol{x}')G_0(\boldsymbol{x}'|\boldsymbol{s}) + G_0(\boldsymbol{r}|\boldsymbol{x}')\varphi(\boldsymbol{x}'|\boldsymbol{s}))f(\boldsymbol{s})\mathrm{d}\boldsymbol{x}', \tag{8}$$

which explains the sensitivity of the reflection data to the change of the background model located at $\boldsymbol{x}'$. Note that, on the right-hand side of equation (8), there two terms in the bracket of the integration, representing the different branches of the reflection wavepaths. As shown in Fig. 1a, the first term denotes the wavefield emitting from the source is first disturbed by a background velocity perturbation, and then reflected by the reflector. The second term denotes the wavefield is first reflected by the reflectivity and then disturbed by the background velocity perturbation (Fig. 1b). For abbreviation, we define the waveform Fréchet kernel as:

$$J = \frac{\partial \Psi}{\partial v_0}. \tag{9}$$

2.2. The reflection traveltime sensitivity kernel

It is well known that the traveltime is more linearly related to the background velocity structure than waveform. Usually, the traveltime misfit is estimated by the cross-correlation between the observed and synthetic data:

$$\tau(\boldsymbol{r}|\boldsymbol{s}) = \underset{\Delta t}{\operatorname{argmax}} \int w_i(t) d^{\mathrm{o}}(\boldsymbol{r}|\boldsymbol{s}, t + \Delta t)\Psi(\boldsymbol{r}|\boldsymbol{s}, t)\mathrm{d}t, \tag{10}$$

where d^{o} is the observed data and $w_i(t)$ is the i-th time window to locally capture the traveltime for a specific event in case of multiple arrivals. The partial derivative of τ with respect to the change of the background velocity can be expressed as:

$$\frac{\partial \tau}{\partial v_0} = \frac{1}{E} \int w_i \dot{d}^{\mathrm{o}}(\boldsymbol{r}|\boldsymbol{s}, t + \tau) \frac{\partial \Psi(\boldsymbol{r}|\boldsymbol{s}, t)}{\partial v_0} \mathrm{d}t, \tag{11}$$

where $E = \int w_i(t)\Psi \ddot{d}^{\mathrm{o}}\mathrm{d}t$ is a normalization factor and the hat dot denotes the time derivative. Assuming that the background model perturbation is very small, then the perturbed data can be taken as a time-shifted version of the true reflection data (Luo and Schuster, 1991), i.e. $d^{\mathrm{o}}(\boldsymbol{r}|\boldsymbol{s}, t + \Delta t) \approx \Psi(\boldsymbol{r}|\boldsymbol{s}, t)$. Therefore, equation (11) can be rewritten as:

$$J_t = \frac{\partial \tau}{\partial v_0} = \int S_i(t) J \mathrm{d}t, \tag{12}$$

where $S_i(t)$ is the weighting factor:

$$S_i(\boldsymbol{r}|\boldsymbol{s}, t) = \frac{w_i(t)\dot{\Psi}(\boldsymbol{r}|\boldsymbol{s}, t)}{\int w_i(t)\Psi(\boldsymbol{r}|\boldsymbol{s}, t)\ddot{\Psi}(\boldsymbol{r}|\boldsymbol{s}, t)\mathrm{d}t}. \tag{13}$$

In this way, we can link the traveltime perturbation to the background velocity perturbations. For a given time window, the time integration in equation (11) or (12) is included during the cross-correlation. Accordingly, the frequency bands of the data will not affect the size of Fréchet derivative because of the cross-correlation. Thus, the size of J_t will be greatly reduced by folding the time or frequency axis, which allows the explicit storage of traveltime Fréchet derivatives if the number of time window is small.

2.3. The inverse problems solved with second-order optimization

The reflection-based inversion aims to retrieve the subsurface model through minimizing the misfit between the observed and simulated reflection data. For instance, RWI updates the background model by solving the following nonlinear optimization problem in a least-squares manner:

$$\chi(v_0) = \frac{1}{2} \sum_{\boldsymbol{r},\boldsymbol{s}} (\boldsymbol{r}|\boldsymbol{s}) - d^{\mathrm{o}}(\boldsymbol{r}|\boldsymbol{s})_2^2. \tag{14}$$

From a numerical point of view, RWI is a large-scale nonlinear inverse problem. We choose the local optimization methods that involve first- and second-order derivatives of the objective function to iteratively solve it. Hence a sequence is v_0^k computed from an initial guess, i.e.,

$$v_0^{k+1} = v_0^k + \alpha^k (\delta v_0)^k, \tag{15}$$

where $(\delta v_0)^k$ represents the model update direction, and α^k is a step-length that can be computed through a line-search procedure.

The first-order derivative of the objective function (i.e., the gradient) can be expressed as:

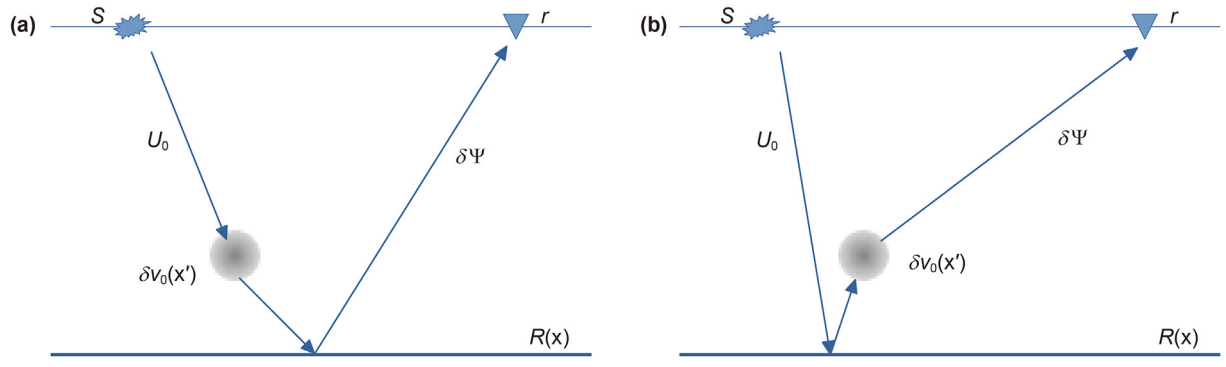

Fig. 1. The schematic illustration of two reflection branches: (**a**) $\boldsymbol{s} \rightarrow \delta v_0(\boldsymbol{x}') \rightarrow R(\boldsymbol{x}) \rightarrow \boldsymbol{r}$, (**b**).$\boldsymbol{s} \rightarrow R(\boldsymbol{x}) \rightarrow \delta v_0(\boldsymbol{x}') \rightarrow \boldsymbol{r}$

$$\nabla\chi = \frac{\partial\chi}{\partial v_0} = J^{\dagger}\delta d, \tag{16}$$

where $\delta d = \Psi(\boldsymbol{r}|\boldsymbol{s}) - d^{o}(\boldsymbol{r}|\boldsymbol{s})$ represents the data residuals. The waveform data are nonlinearly related to the low-to-intermediate wavenumber of the velocity model. In the framework of Born modelling, the accuracy of the simulated data depends on the appropriacy of the background and perturbation models. If there is an obvious lack of low wavenumbers in the starting model, the estimated perturbation model will be far away from the true one, and the simulated reflection data have evident kinematic errors. In this setting, RWI tends to converge to a local minimum due to cycle skipping problem. To tackle this challenge, a common strategy is to start with the traveltime inversion because traveltime is more linearly related to the background velocity (Ma and Hale, 2013; Chi et al., 2015; Brossier et al., 2015; Xu et al., 2019).

In the context of RTI, we build the objective function with reflection traveltime difference, namely:

$$\chi_t(v_0) = \frac{1}{2}\sum_{\boldsymbol{r},\boldsymbol{s}}\tau(\boldsymbol{r}|\boldsymbol{s})_2^2. \tag{17}$$

Accordingly, the gradient with respect to the background model is changed as:

$$\nabla\chi_t = \frac{\partial\chi_t}{\partial v_0} = J_t^{\mathrm{T}}\tau, \tag{18}$$

where T denotes the transpose. The widely used first-order optimization methods, such as the steep-descent and nonlinear CG methods, only involves the calculation of the gradient. In the context of RWI or RTI, this can be efficiently implemented by using the adjoint-state method (e.g., Xu et al., 2012; Ma and Hale, 2013).

Taking into account the second-order derivative of the objective function (namely the Hessian) and only keeping the linear term, we have the Gauss-Newton equation for RWI and RTI:

$$H\delta v_0 = -\nabla\chi,\ H_t\delta v_0 = -\nabla\chi_t, \tag{19}$$

in which the approximate Hessian matrices are constructed by cross-correlation of the Fréchet derivative wavefields:

$$H = J^{\dagger}J, \quad H_t = J_t^{\mathrm{T}}J_t. \tag{20}$$

At each nonlinear RWI or RTI iteration, either the second-order adjoint-state method or the improved scattering-integral method is used to solve the linear system 18, as in the context of Gauss-Newton FWI (Clement et al., 2001; Plessix, 2006; Metivier et al., 2014; Liu et al., 2015; Pan et al., 2016). These matrix-free CG methods avoid to explicitly form the Hessian matrix and only the Hessian-vector product is required at each iteration of the CG algorithm. The readers are referred to literature, e.g., Metivier et al. (2014); Liu et al. (2015), for the technical details.

3. Practical strategy for large-scale applications

The application of the second-order optimization method to large-scale RTI or RWI problem faces great challenges. First, the inversion generally requires careful preconditioning to get a reasonable model update due to the ill-posedness of this kind of inverse problem. Second, extensive computational resources are required even for a 2-D real data case. The appropriate strategies, such as the decimation in the data or model domain, have to be taken into consideration to make the Gauss-Newton method feasible for the real data applications, although it may sometimes sacrifice the accuracy. Here, we will design some appropriate preconditioning and computational strategies to balance the accuracy and efficiency for the reflection-based inversion.

3.1. Precondition the Hessian matrix

In a layered medium with many interfaces at different depths, the Hessian matrix of RWI may behave very complicated. First, the reflections at different depths and of different offsets experience different geometric spreading during wavefield propagation. Second, the magnitudes of the reflectivity in Born modeling significantly affects the synthetic reflection amplitudes. Therefore, the magnitudes of the Fréchet kernels for different reflection events could varies in a very wide range. As a result, the elements of Hessian matrix corresponding to the deep locations generally have very small magnitudes. To illustrate this, we use a toy model with two reflectors for the RWI Hessian calculation, which is shown in Fig. 2.

We observe that the elements for the first layer are much larger than those of the second layer in the RWI Hessian matrix (Fig. 2b). This kind of imbalance increases the ill-conditioning and may lead to unreasonable model updating in the deeper part, since the very small values in the Hessian matrix are truncated during the inner iterations of the Gauss-Newton method. As Métivier et al. (2017) discussed in the context of FWI, the truncated Gauss-Newton method will benefit from introducing an appropriate preconditioner for the inner linear system. Here we choose to precondition the Hessian in the presence of many interfaces through normalizing the forward modeled data as well as the reflectivity to balance the Fréchet derivative wavefields at different depths (see Fig. 2c). In the numerical examples, we find that this preconditioning improves the Gauss-Newton RTI and RWI algorithms, although it may sometimes sacrifice the resolving power a little bit.

3.2. The computational optimization

The second-order methods using the Gauss-Newton Hessian can achieve a locally super-linear convergence rate. However, the significantly increased computational burden to solve the inner linear system using the CG method is a major challenge in practice. In the context of FWI, gradients and Hessian-vector products that usually calculated through the adjoint-state method (Metivier et al., 2014) or the scattering integral (SI) method (Chen et al., 2007), require to perform a huge number of forward modeling. The number is linearly proportional to the number of receivers or sources in the data. In the context of RWI, if using the SI method to solve the Gauss-Newton equation, the Fréchet derivative should be calculated and stored explicitly for each source-receiver pair. As indicated by equation (8), we need two wavefields of the source side ($G_0(\boldsymbol{x}'|\boldsymbol{s})$, $\varphi(\boldsymbol{x}'|\boldsymbol{s})$) and two wavefields of the receiver side ($G_0(\boldsymbol{x}'|\boldsymbol{r})$, $\varphi(\boldsymbol{x}'|\boldsymbol{r})$), making the total number of forward modeling $2N_s + 2N_r$ for each outer loop, where N_s and N_r are the number of source and receiver, respectively. And the size of J will be as large as $N_{sr} \times M$ for a single frequency, with N_{sr} the total number of source-receiver pairs and M the model size for each shot. Accordingly, the Gauss Newton RWI using SI method requires the huge amount of computational cost and storage consumption, even for the 2D problems.

In order to make the SI-based Gauss Newton RWI feasible, we introduce the following strategies to lower the computational burden. First, we choose to use the coarse grids to store the Fréchet derivative wavefields, because RTI and RWI aim to recover the low-to-intermediate wavenumbers and thus a coarse-grid discretization is sufficient to represent the model. With the frequencies increasing, we synthesize the reflection wavefield with finer grids but still store the Fréchet kernels with the coarse grids. Second, we

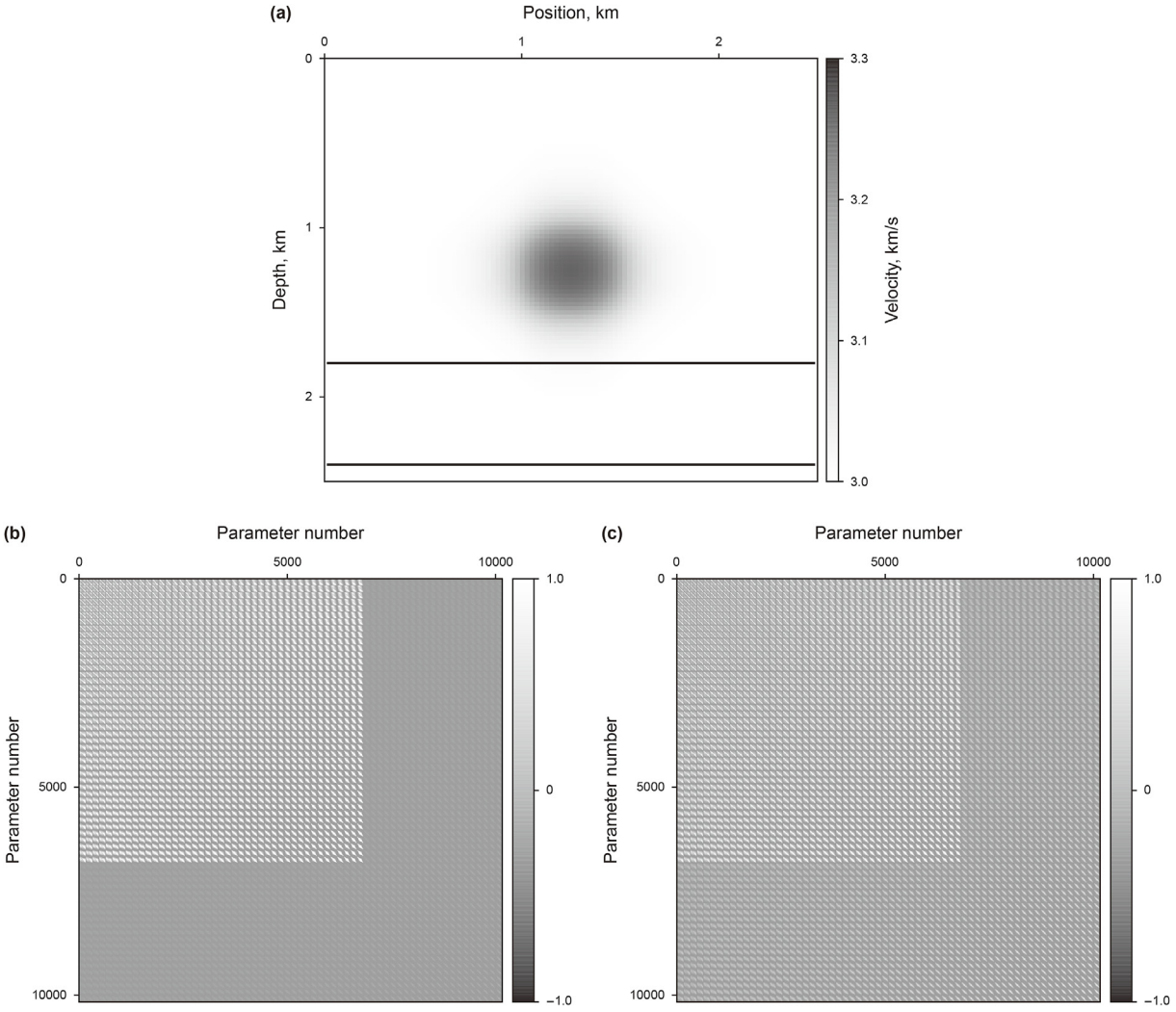

Fig. 2. Precondition the RWI Hessian: (**a**) the modified toy model with two interfaces of the same reflectivity at the depth of 1.8 km and 2.3 km; (**b**) the original and (**c**) preconditioned Hessian matrices.

decimate the receivers to calculate and save the Fréchet kernels. In a certain shot gather, the neighboring receivers possess quite similar reflection wave-paths, which actually implies the redundant information for the inversion. By sharing a same kernel for neighboring traces within a receiver group, we can significantly reduce the number of wave equation solving and also save the disk overhead to store the Fréchet derivatives. As shown in Fig. 3, if we change from standard grid $d_x \times d_z$ to the coarse grid $D_x \times D_z$ and group the neighboring receivers as one trace (the receiver number inside the group is $N_r^{'}$), then the storage will reduce $\frac{D_x D_z N_r^{'}}{d_x d_z}$ times, and the total number of forward modeling will reduce $N_r^{'}$ times. In addition, the frequencies are also redundant to satisfy the wavenumber coverage and can be decimated into several discrete groups (Sirgue and Pratt, 2004; Brenders and Pratt, 2007). Accordingly, for the RWI problem, we can further reduce the size of Fréchet derivative by only storing several frequency slices during the hierarchical inversion from low to high frequencies. Note that, the sizes of coarse-grid and receiver group should be in accordance with the frequency used in the certain stage.

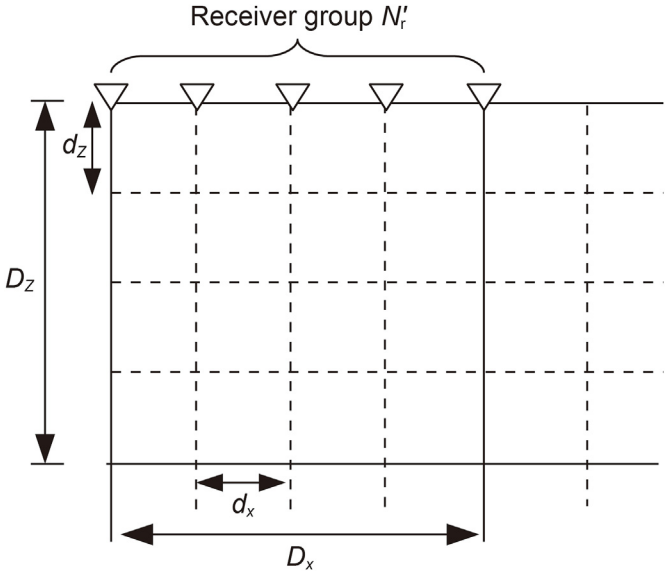

Fig. 3. Schematic illustration of the computational strategies. The Fréchet kernel will be stored in the coarse grid ($D_x \times D_z$), while the forward modeling is applied on the fine grid ($d_x \times d_z$). In the meantime, the neighboring receivers will share one Frechet kernel to save the computational and storage overburden.

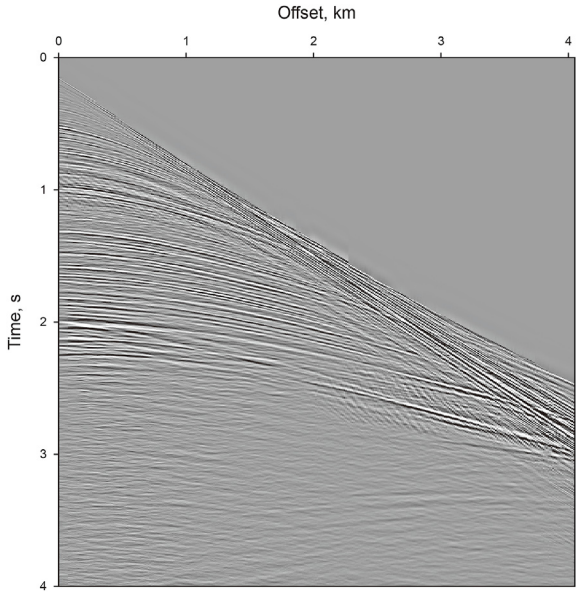

Fig. 4. One common-shot gather of the East China Sea dataset.

4. Field data example

Based on the above strategies, we propose a practical two-stage workflow of reflection-based inversion using the Gauss-Newton method to reconstruct the velocity model. First, we apply Gauss-Newton RTI to reconstruct the low-wavenumber parts of the velocity structures. Then, Gauss-Newton RWI is followed to gradually supplement more intermediate wavenumbers to the model. We test the proposed workflow with a real data from East China Sea.

The 2D dataset of 851 shots was acquired by using air guns and towed streamers in East China Sea. The shot interval is 37.5 m. The spacing of the hydrophones is 12.5 m and the maximum offset is 4 km. Fig. 4 displays one common-shot profile. Conventional data processing, including noise attenuation, designature, deghost and demultiple, have been applied before the inversion. The legacy model for migration is converted from the time-domain root mean square (RMS) velocity, which roughly represents the background features of the target area. However, it is inadequate to correctly honor the kinematics for RTM to provide well-focused migration image and flatten common-image gathers (CIGs) (see Fig. 5).

We apply the proposed two-stage reflection-based inversion to improve the velocity model. The hierarchical strategy with frequency continuation is implemented for both GN-RTI and GN-RWI. In the RTI stage, the resolution of traveltime estimation will be greatly affected by the frequency bands. Therefore, we select three frequency bands, namely 4–8 Hz, 4–12 Hz and 4–16 Hz, to gradually increase the resolution of traveltime estimation along with adaptively designed sliding windows. While in the RWI stage, three frequency groups, 4–8 Hz, 8–12 Hz and 8–16 Hz with an increment of 1 Hz for each group, are used for inversion. For both RTI and RWI, we set 10 outer loops with a maximum 5 iterations in the inner loop for each frequency groups. In the inner loop of Gauss-Newton algorithm, we precondition the Hessian to guarantee a robust convergence of inversion. For comparison, we also perform the conventional workflow based on RTI and RWI with CG method (i.e., CG-RTI and CG-RWI) to show the effectiveness of the proposed Gauss-Newton algorithm. In addition, the structure-oriented regularization (Yu et al., 2020) are utilized in the aforementioned inversion process to make velocity update geologically consistent.

First, we apply the conventional workflow based on CG-RTI and CG-RWI, whose results are shown in Fig. 6. We observe that the velocity update from CG-RTI mainly relates to the strong interfaces, which is not able to flat the CIGs. Starting from CG-RTI model, the velocity is improved after the CG-RWI and thus CIGs are significantly flattened. However, we still observe some residual moveouts

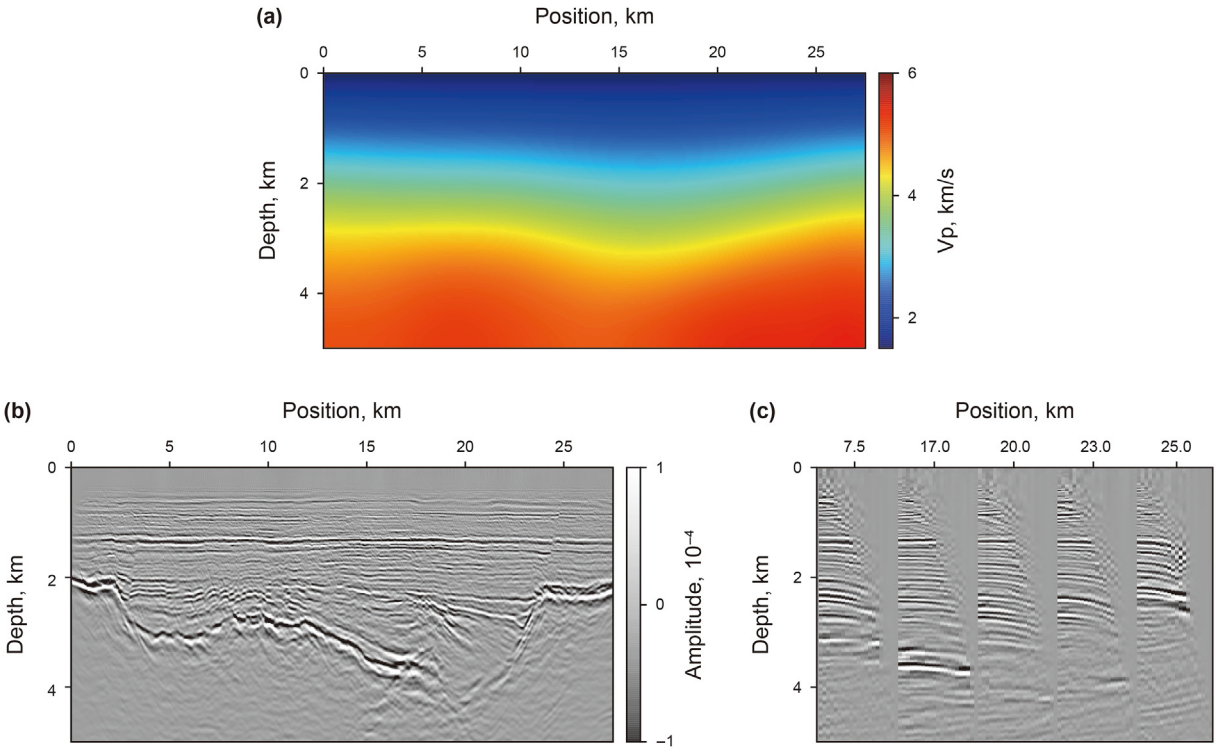

Fig. 5. Migration results with initial model: (**a**) the initial velocity model, (**b**) RTM image and (**c**) Offset-domain CIGs.

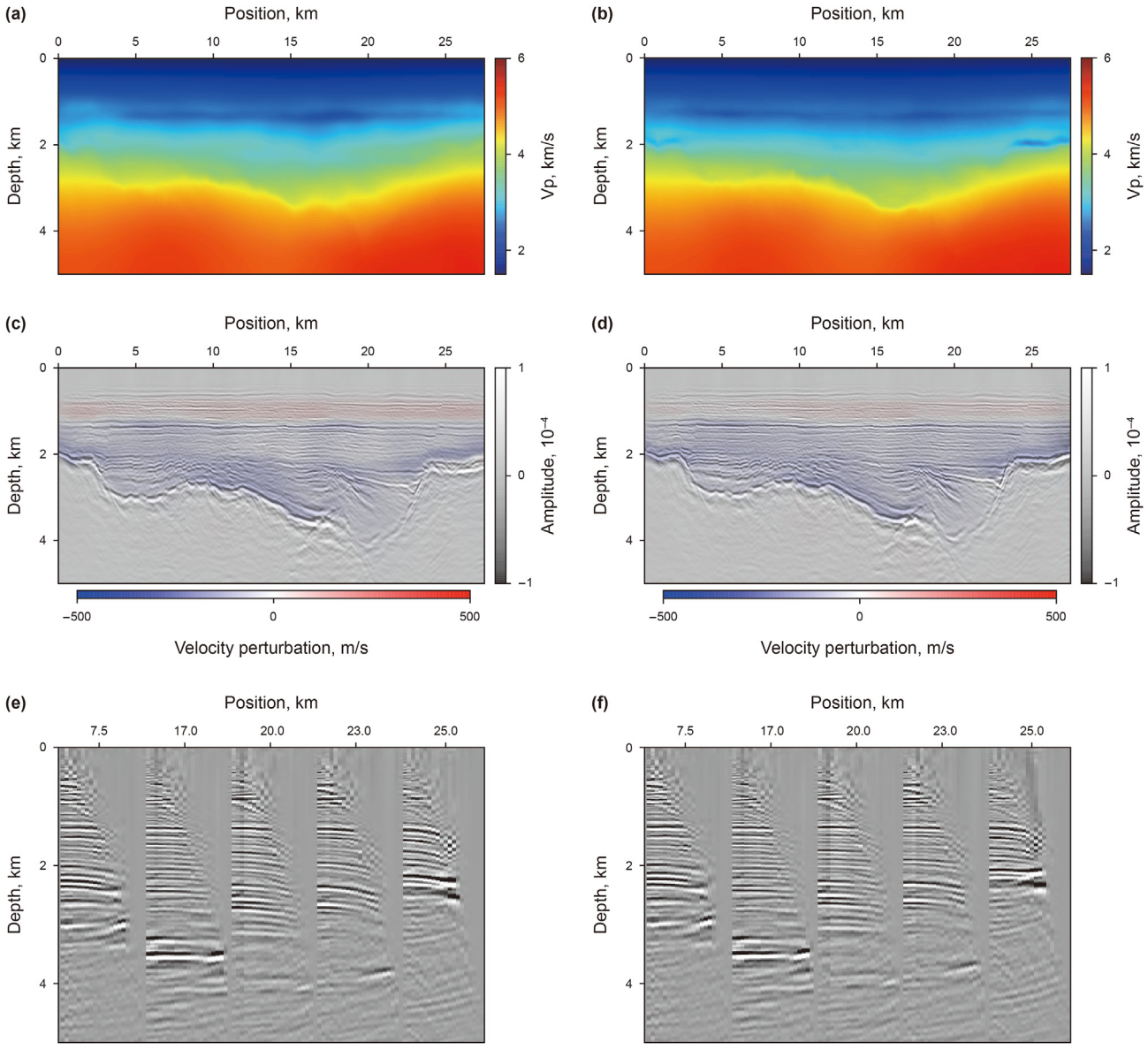

Fig. 6. The results from conventional workflow with CG-RTI and CG-RWI. (**a**) and (**b**): the CG-RTI and CG-RWI velocity model. (**c**) and (**d**): the RTM image using (**a**) and (**b**) overlaid by the total velocity update. (**e**) and (**f**): the offset-domain CIGs using (**a**) and (**b**).

on the CIGs, which implies that although the conventional RTI and RWI can provide reasonable migration velocity model, the accuracy of velocity still remains to be improved. For comparison, we use the proposed Gauss-Newton based workflow to update the velocity. After the Gauss-Newton RTI, the velocity is obviously updated for both the shallow and the deep parts (Fig. 7c). The basement imaging of the sedimentary basin is greatly improved with better continuity. The deep faults at the position of 20 km becomes clearer. However, since the multi-window cross-correlation may fail to accurately measure the relatively small traveltime residuals in the latter iterations of this stage, there are still some high-wavenumber image footprints on the velocity and the CIGs also exhibit some residual moveouts. Then, starting from the RTI model, we apply the Gauss-Newton RWI, which is sensitive to the waveform misfit, to further improve the recovery of intermediate wavenumbers of the velocity model. As shown in Fig. 7b, after the GN-RWI, the velocity anomaly along the strong reflection interface disappears. Some adjustments of velocity model can be observed below the depth of 1.5 km (Fig. 7d), which essentially honors the correct kinematics to flatten the CIGs. The final migrated image based on GN-RWI model shows more continuous sequence boundaries and clear stratigraphic textures.

5. Discussion

In the stage of RTI, the cross-correlation only estimate the traveltime residuals of the dominant event inside the selected time window. In the presence of multi-events with various intensities, the cross-correlation associated with running windows may fail to provide high-resolution measurement of the traveltime misfits. Other methods, such as dynamic image warping (Hale, 2013), frequency-dependent traveltime measurement (Laske and Masters, 1996), and instantaneous phase (Bozdag et al., 2011; Choi et al., 2013) and so on, all have their own disadvantages in this situation. This means that the reflection traveltime or phase inversion has limitation in spatial resolution. That is why the waveform

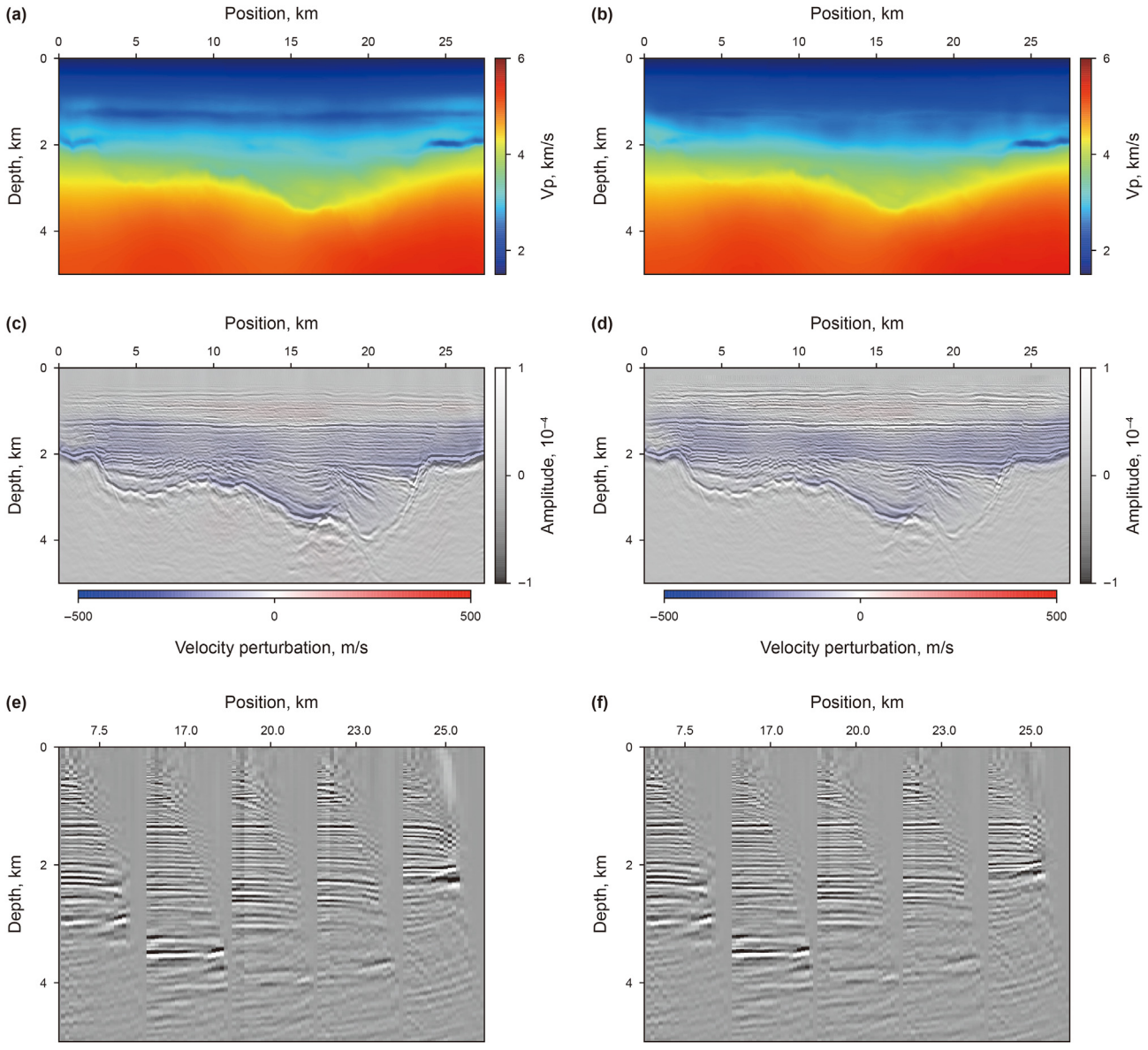

Fig. 7. The results from proposed workflow with GN-RTI and GN-RWI. (**a**) and (**b**): the GN-RTI and GN-RWI velocity model. (**c**) and (**d**): the RTM image using (**a**) and (**b**) overlaid by the total velocity update. (**e**) and (**f**): the offset-domain CIGs using (**a**) and (**b**).

inversion is followed to recover the model intermediate-wavenumbers. However, waveform fitting highly depends on the accuracy of the estimated reflectivity (or high-wavenumber model perturbations) and seismic wave propagation physics. According to the radiation pattern, e.g., Wang and Cheng (2017), the density perturbation has a remarkable coupling effect with the P-wave velocity perturbation at small scattering angles. This implies that the effect of density needs to be taken into account in estimation of the high-wavenumber model perturbations to improve the amplitude accuracy of the simulated reflection data. For instance, least-squares RTM based on the variable-density acoustic wave propagator (Yang et al., 2016) could be used for this purpose.

The adjoint-state method plays an important role in the seismic waveform tomography, through which the functional gradient can be efficiently calculated by cross-correlating the forward propagated wavefields with the back-propagated adjoint wavefields (Plessix, 2006; Virieux and Operto, 2009; Fichtner and Trampert, 2011). In the context of FWI, both the first- and second-order adjoint-state method are closely related to the scattering-integral method (Tromp et al., 2005; Chen et al., 2007; Epanomeritakis et al., 2008; Liu et al., 2015), but which one is better relies on the acquisition geometry, particularly on the ratio of sources to receivers, as well as the trade-offs between computing cost and file input/output operations (Chen et al., 2007). In this paper, we employ the scattering-integral to calculate the Hessian-vector product in the nested inner loops of Gauss-Newton RWI or RTI. However, for the modern marine and land acquisition system in seismic exploration, the number of receivers is far more than the number of sources, which means that the calculation and storage of Fréchet derivatives will be extremely resource-demanding without proper computational strategies as aforementioned in the previous sections. Therefore, the second-order adjoint-state method for Gauss-Newton RWI or RTI, which can calculate the Hessian-vector product without explicitly storing the Fréchet derivatives (Fichtner and Tramper, 2011; Metivier et al., 2017), will be an appropriate alternative in the real data applications (Wang et al., 2021).

6. Conclusions

To improve velocity model building using precritical seismic reflection data, we have reviewed the reflection waveform and traveltime Fréchet derivatives, and demonstrated Hessian-based second-order optimization methods in the context of RWI and RTI. Accordingly, we have proposed a practical workflow cascading Gauss-Newton RTI and Gauss-Newton RWI, in which Hessian preconditioning and computational strategies, such as coarse grid sampling, receiver decimation and frequency grouping are suggested to make the workflow robust and efficient. The application to a towed-streamer data set with limited offsets has demonstrated that the two-stage workflow is qualified to improve the velocity model and structural imaging in the deep part of the sedimentary basin.

Acknowledgements

This work is supported by National Natural Science Foundation of China (42074157), the National Key Research and Development Program of China (2018YFC0310104) and the Strategic Priority Research Program of the Chinese Academy of Science (XDA14010203).

References

Adler, F., Baina, R., Soudani, M.A., et al., 2008. Non-linear 3d tomographic least-squares inversion of. residual moveout in Kirchhoff prestack-depth-migration common-image gathers. Geophysics 73, VE13–VE23. https://doi.org/10.1190/1.2956427.

Bozdag, E., Trampert, J., Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. Geophys. J. Int. 185, 845–870. https://doi.org/10.1111/j.1365-246X.2011.04970.x.

Brenders, A., Pratt, R., 2007. Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model. Geophys. J. Int. 168, 133–151. https://doi.org/10.1111/j.1365-246X.2006.03156.x.

Brossier, R., Operto, S., Virieux, J., 2009. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. Geophysics 74 (6), WCC105–WCC118. https://doi.org/10.1190/1.3215771.

Brossier, R., Operto, S., Virieux, J., 2015. Velocity model building from seismic reflection data by full waveform inversion. Geophys. Prospect. 63, 354–367. https://doi.org/10.1111/1365-2478.12190.

Chen, P., Jordan, T.H., Zhao, L., 2007. Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods. Geophys. J. Int. 170, 175–181. https://doi.org/10.1111/j.1365-246X.2007.03429.x.

Chi, B.X., Dong, L.G., Liu, Y.Z., 2015. Correlation based reflection full waveform inversion. Geophysics 80, R189–R202. https://doi.org/10.1190/geo2014-0345.1.

Choi, Y., Alkhalifah, T., 2013. Frequency-domain waveform inversion using the phase derivative. Geophys. J. Int. 195, 1904–1916. https://doi.org/10.1093/gji/ggt351.

Claerbout, J.F., 1971. Toward a unified theory of reflector mapping. Geophysics 36, 467–481. https://doi.org/10.1190/1.1440185.

Claerbout, J.F., 1985. Imaging the Earth's Interior. Blackwell Scientific. https://doi.org/10.1111/j.1365-246X.1986.tb01086.x.

Clement, F., Chavent, G., Gomez, S., 2001. Migration-based traveltime waveform inversion of 2d simple structures: a synthetic example. Geophysics 66, 845–860. https://doi.org/10.1190/1.1444974.

Epanomeritakis, I., Akcelik, V., Ghattas, O., et al., 2008. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion. Inverse Probl. 24, 034015. https://doi.org/10.1088/0266-5611/24/3/034015.

Fichtner, A., Trampert, J., 2011. Hessian kernels of seismic data functionals based upon adjoint techniques. Geophys. J. Int. 185, 775–798. https://doi.org/10.1111/j.1365-246X.2011.04966.x.

Guillaume, P., Lambare, G., Leblanc, O., et al., 2008. Kinematic. invariants: an efficient and approach for velocity model building. SEG Tech. Progr. Expand. Abstr. https://doi.org/10.1190/1.3064100.

Hale, D., 2013. Dynamic warping of seismic images. Geophysics 78, S105–S115. https://doi.org/10.1190/geo2012-0327.1.

Innanen, K., 2014. Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion. Geophys. J. Int. 199 (2), 717–734. https://doi.org/10.1093/gji/ggu291.

Laske, G., Masters, G., 1996. Constraints on global phase velocity maps from long-period polarization data. J. Geophys. Res. Solid Earth 101, 16059–16075. https://doi.org/10.1029/96JB00526.

Liu, Y.Z., Yang, J.Z., Chi, B.X., et al., 2015. An improved scattering-integral approach for frequency-domain full waveform inversion. Geophys. J. Int. 202, 1827–1842. https://doi.org/10.1093/gji/ggv254.

Luo, Y, Schuster, G.T., 1991. Wave-equation traveltime inversion. Geophysics 56 (5), 645–653. https://doi.org/10.1190/1.1443081.

Ma, Y., Hale, D., 2013. Wave-equation reflection traveltime inversion with dynamic warping and full waveform inversion. Geophysics 78, R223–R233. https://doi.org/10.1190/geo2013-0004.1.

Metivier, L., Bretaudeau, F., Brossier, R., et al., 2014. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures. Geophys. Prospect. 62, 1353–1375. https://doi.org/10.1111/1365-2478.12136.

Metivier, L., Brossier, R., Operto, S., et al., 2017. Full waveform inversion and the truncated Newton method. SIAM Rev. 59, 153–195. https://doi.org/10.1137/16M1093239.

Operto, S., Gholami, Y., Prieux, V., et al., 2013. A guided tour of multiparameter full-waveform inversion with multicomponent data: from theory to practice. Lead. Edge 32, 1040–1054. https://doi.org/10.1190/tle32091040.1.

Pan, W.Y., Innanen, K.A., Liao, W.Y., 2017. Accelerating Hessian-free Gauss-Newton full-waveform inversion via l-BFGS preconditioned conjugate-gradient algorithm. Geophysics 82 (2), R49–R64. https://doi.org/10.1190/geo2015-0595.1.

Pan, W.Y., Innanen, K.A., Margrave, G.F., et al., 2016. Estimation of elastic constants for HTI media using gauss-Newton and full Newton multiparameter full-waveform inversion. Geophysics 81, R275–R291. https://doi.org/10.1190/geo2015-0594.1.

Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. Geophys. J. Int. 167, 495–503. https://doi.org/10.1111/j.1365-246X.2006.02978.x.

Pratt, R.G., Shin, C., Hick, G., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. Geophys. J. Int. 133, 341–362. https://doi.org/10.1046/j.1365-246X.1998.00498.x.

Shin, C., Jang, S., Min, D.J., 2001. Improved amplitude preservation for prestack depth migration by inverse scattering theory. Geophys. Prospect. 49 (5), 592–606. https://doi.org/10.1046/j.1365-2478.2001.00279.x.

Sirgue, L., Pratt, R.G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal. frequencies. Geophysics 69, 231–248. https://doi.org/10.1190/1.1649391.

Stork, C., 1992. Reflection tomography in the postmigrated domain. Geophysics 57, 680–692. https://doi.org/10.1190/1.1443282.

Tromp, J., Tape, C., Liu, Q.C., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. Geophys. J. Int. 160, 195–216. https://doi.org/10.1111/j.1365-246X.2004.02453.x.

Virieux, J., Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics. Geophysics 74, WCC1–WCC26. https://doi.org/10.1190/1.3238367.

Wang, T.F., Cheng, J.B., 2017. Elastic full waveform inversion based on mode decomposition: the approach and mechanism. Geophys. J. Int. 209, 606–622. https://doi.org/10.1093/gji/ggx038.

Wang, T.F., Cheng, J.B., Guo, Q., et al., 2018. Elastic wave-equation-based reflection kernel analysis and. traveltime inversion using wave mode decomposition. Geophys. J. Int. 215, 450–470. https://doi.org/10.1093/gji/ggy291.

Wang, H.Y., Singh, S.C., Audebert, F., et al., 2015. Inversion of seismic refraction and reflection data for building long-wavelength velocity models. Geophysics 80 (2), R81–R93. https://doi.org/10.1190/geo2014-0174.1.

Wang, T.F., Xu, W.C., Cheng, J.B., et al., 2020. Practical Reflection Waveform Inversion Using Truncated Gauss-Newton Method. 90th Annual International Meeting SEG Expanded Abstracts, pp. 835–839. https://doi.org/10.1190/segam2020-3427176.1.

Wang, T.F., Cheng, J.B., Geng, J.H., 2021. Reflection full waveform inversion with second-order optimization using the adjoint-state method. J. Geophys. Res. Solid Earth 126, e2021JB022135. https://doi.org/10.1029/2021JB022135.

Wang, Y.W., Dong, L.G., Liu, Y.Z., et al., 2016. 2D frequency-domain elastic full-waveform inversion using the block-diagonal pseudo-Hessian approximation. Geophysics 81 (5), R429–R445. https://doi.org/10.1190/geo2015-0678.1.

Wang, Y., Pratt, R.G., 1997. Sensitivities of seismic traveltimes and amplitudes in reflection tomography. Geophys. J. Int. 131, 618–642. https://doi.org/10.1111/j.1365-246X.1997.tb06603.x.

Woodward, M.J., 1992. Wave-equation tomography. Geophysics 57, 15–26. https://doi.org/10.1190/1.1443179.

Woodward, M.J., Nichols, D., Zdraveva, O., et al., 2008. A decade of tomography. Geophysics 73. https://doi.org/10.1190/1.2969907. VE5-VE11.

Wu, Z.D., Alkhalifah, T., 2015. Simultaneous inversion of the background velocity and the perturbation in full-waveform inversion. Geophysics 80, R317–R329. https://doi.org/10.1190/geo2014-0365.1.

Xu, S., Wang, D., Chen, F., et al., 2012. Inversion on reflected seismic wave. 82nd SEG Expanded Abstracts. https://doi.org/10.1190/segam2012-1473.1.

Xu, W.C., Wang, T.F., Cheng, J.B., 2019. Elastic model low-to-intermediate-wavenumber inversion using reflection traveltime and waveform of multi-component seismic data. Geophysics 84, R109–R123. https://doi.org/10.1190/geo2018-0306.1.

Yang, J.Z., Liu, Y.Z., Dong, L.G., 2016. Least-squares reverse time migration in the presence of density variations. Geophysics 81, S497–S509. https://doi.org/

10.1190/GEO2019-0005.1.
Yao, G., Wu, D., 2017. Reflection full waveform inversion. Sci. China Earth Sci. 60, 17831794. https://doi.org/10.1007/s11430-016-9091-9.
Yao, G., Wu, D., Wang, S.X., 2020. A review on reflection-waveform inversion. Petrol. Sci. 17, 334–351. https://doi.org/10.1007/s12182-020-00431-3.
Zhang, Y., Ratcliffe, A., Roberts, G., et al., 2014. Amplitude-preserving reverse time migration: from reflectivity to velocity and impedance inversion. Geophysics 79, S271–S283. https://doi.org/10.1190/geo2013-0460.1.
Zhou, W., Brossier, R., Operto, S., et al., 2015. Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation. Geophys. J. Int. 202, 1535–1554. https://doi.org/10.1093/gji/ggv228.