Petroleum Science 23 (2026) 4644–4661

Contents lists available at ScienceDirect

Petroleum Science

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

Original Paper

Hessian-based elastic reflection waveform inversion with scale fusion strategy

Ming-Qian Wang [a], Bing-Shou He [a,b,*]

[a] Key Laboratory of Submarine Geosciences and Prospecting Techniques, Ministry of Education of the People's Republic of China, College of Marine Geosciences, Ocean University of China, Qingdao, 266100, Shandong, China

[b] Evaluation and Detection Technology Laboratory of Marine Mineral Resources, Laoshan National Laboratory of Marine Science and Technology, Qingdao, 266237, Shandong, China

A R T I C L E I N F O

Article history:
Received 21 February 2025
Received in revised form
22 December 2025
Accepted 30 March 2026
Available online 4 April 2026

Edited by Meng-Jiao Zhou

Keywords:
Seismic waveform inversion
Inverse problems
Hessian matrix
Elastic wave equation

A B S T R A C T

Full waveform inversion (FWI) theoretically has the potential to recover high-resolution and accurate velocity models. However, FWI primarily relies on refractions and turning waves, and therefore often fails to update the deep background velocity under limited-offset acquisition geometries. Reflection waveform inversion (RWI) can recover the low-to intermediate-wavenumber components of the velocity model from deeply penetrating reflected waves, thereby improving the background velocity model. Elastic reflection waveform inversion (ERWI) exploits multicomponent seismic data to simultaneously invert for P-wave and S-wave velocities, providing enriched information about subsurface structures. The construction of perturbation models plays a critical role in ERWI, as these models not only influence the background model updating process but also constitute an integral part of the final inversion results. However, conventional ERWI methods rely on gradient-based optimization and neglect the contribution of the Hessian matrix, leading to lower accuracy of the perturbation models. The Hessian matrix accounts for the effects of amplitude imbalance, uneven illumination, band-limited wavelets, and multi-parameter trade-offs. To overcome these limitations, we propose a truncated Gauss-Newton ERWI method in which the perturbation models are updated using efficient Hessian-vector products and a matrix-free conjugate gradient algorithm. Furthermore, based on the analysis of the Fréchet derivatives and gradient expressions, we introduce a scale-fusion strategy for ERWI. This strategy enables effective utilization of the perturbation models while progressively improving their accuracy throughout the iterative inversion process, thereby establishing a mutually reinforcing feedback mechanism between the perturbation and background models. Numerical experiments demonstrate that the proposed method achieves faster convergence, lower data misfit and model errors, and significantly improved inversion quality, while exhibiting strong robustness to noise.

1. Introduction

Full waveform inversion (FWI) is a powerful seismic inversion technique used to obtain high-resolution subsurface models by minimizing the misfit between observed and predicted seismic data (Lailly and Bednar, 1983; Tarantola, 1984; Virieux and Operto, 2009). Unlike traditional methods that rely on ray-based approaches or travel-time tomography, FWI exploits the entire seismic waveform, including both amplitude and phase information, enabling the retrieval of detailed subsurface properties (Operto et al., 2006; Brenders and Pratt, 2007; Sirgue et al., 2010). With the continuous improvement in computational power and the efficient implementation of the adjoint-state method (Plessix, 2006), FWI has been widely applied to reconstruct regional subsurface models and oil and gas exploration (Sirgue and Pratt, 2004; Prieux et al., 2013; Vigh et al., 2014; Zhang et al., 2016; Wang and Cheng, 2017; Yao et al., 2018; Xie et al., 2024).

Due to the strong nonlinearity between the model parameters and seismic data governed by the wave equation, FWI is prone to being trapped in local minima during the inversion process. Consequently, the successful application of FWI strongly depends on a good initial model and the availability of low-frequency data

* Corresponding author.
E-mail address: hebinshou@ouc.edu.cn (B.-S. He).

Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2026.03.065
1995-8226/

(Bunks et al., 1995). In the absence of low-frequency, an accurate initial model becomes even more critical to avoid cycle skipping and ensure reliable inversion results. To alleviate the dependence of FWI on the initial model and low-frequency data, various strategies have been proposed. One class of methods modifies the conventional L_2-norm objective function to enhance its convexity (Van Leeuwen and Mulder, 2010; Métivier et al., 2016; Yong et al., 2019; Da Silva et al., 2022). Another class extends the model space to improve the inversion process (Shin and Cha, 2008; Ha et al., 2015; Warner and Guasch, 2016; Huang et al., 2018). In addition, envelope-based and multiscale inversion methods have also been developed (Wu et al., 2014; Chi et al., 2014; Bunks et al., 1995). Nevertheless, the recovery of the deep background velocity remains particularly challenging. In conventional FWI formulations, the background velocity is mainly updated using refractions and turning waves, which requires long-offset data to invert the deep part of the model (Yao et al., 2020). When such data are unavailable or limited, conventional FWI often fails to correctly update the deep background velocity, leading to poor convergence and inaccurate model reconstruction.

To overcome these limitations, reflection waveform inversion (RWI) was proposed as an alternative strategy. RWI decomposes the velocity model into a smooth background model and a perturbation model based on scale separation, and recovers the long-wavelength components of the velocity model using reflected waves predicted through migration and demigration processes (Xu et al., 2012; Ma and Hale, 2013; Brossier et al., 2015). The RWI framework involves updating both the perturbation and background models. The perturbation model is typically updated in each iteration through least-squares migration, while the perturbation and background models are updated alternately. The low-to intermediate-wavenumber updates of the background model in RWI rely on the tomographic components of reflected waves. By jointly incorporating incident waves and reflected waves, the inversion performance of RWI can be further improved (Zhou et al., 2015; Wu and Alkhalifah, 2015). Moreover, filtering strategies based on subsurface structure and scattering angles have been proposed to enhance the inversion accuracy of RWI (Alkhalifah, 2014; Yao et al., 2020).

FWI and RWI based on the acoustic approximation have been extensively studied. However, incorporating elastic effects enables the retrieval of richer subsurface information, such as P-wave velocity, S-wave velocity, and density, thereby improving the resolution and accuracy of the reconstructed subsurface models. Nevertheless, multi-parameter elastic FWI (EFWI) suffers from pronounced parameter trade-off effects. Hierarchical inversion strategies and wave-mode decoupling techniques have been introduced to mitigate parameter crosstalk and improve inversion performance (Ren and Liu, 2016; Wang and Cheng, 2017). Elastic reflection waveform inversion (ERWI) has been proposed to recover the low-to intermediate-wavenumber components of P-wave and S-wave velocity models using reflected waves (Guo and Alkhalifah, 2017; Xu et al., 2019; Li et al., 2019).

Traditional first-order derivative-based optimization methods rely solely on gradient information. Although they are straightforward to implement and computationally efficient per iteration, they usually require a large number of iterations to converge and may fail to produce satisfactory inversion results. In contrast, second-order Newton-type optimization methods incorporate the Hessian matrix, which represents the second-order derivative of the objective function with respect to the model parameters. The Hessian matrix accounts for the effects of finite acquisition aperture, geometric spreading, and band-limited wavelets. In multi-parameter inversion problems, it can also mitigate parameter trade-off effects. Hessian-based second-order optimization methods have been successfully applied in FWI to accelerate convergence and improve inversion accuracy (Fichtner and Trampert, 2011; Liu et al., 2015; Yang et al., 2016; Métivier et al., 2017; Pan et al., 2017; Chen and Sacchi, 2020; Eslaminia et al., 2022; Xing and Zhu, 2024). In the context of RWI, Gauss-Newton methods based on the second-order adjoint-state method and scattering integral formulations have also been developed to enhance inversion performance (Xu et al., 2021; Wang et al., 2022). For multi-parameter ERWI, incorporating Hessian information is particularly important due to the presence of strong parameter trade-offs. Introducing the Hessian into ERWI can improve convergence and inversion accuracy.

In this study, we propose a Hessian-based ERWI framework incorporating a scale-fusion strategy to improve the convergence and resolution of conventional ERWI. A joint incident wave and reflected wave objective function is employed to make more effective use of the available data information. Within this framework, the perturbation model at each iteration is temporarily constructed by solving the Newton equation in the inner loop of the Gauss-Newton method, yielding P- and S-wave perturbation models with higher accuracy and reduced multi-parameter crosstalk. These improved perturbation models facilitate more accurate computation of the background model gradients. The proposed scale-fusion strategy is theoretically justified, as the Fréchet derivatives and gradients of the perturbation models naturally appear in the corresponding expressions for the background models. This strategy ensures that the perturbation models are effectively exploited while their accuracy is progressively enhanced during the iterative inversion process, thereby establishing a mutually reinforcing feedback mechanism between the perturbation and background models. Numerical experiments demonstrate that the proposed ERWI method achieves improved convergence, more accurate inversion results, and strong robustness to noise.

2. Methodology

2.1. Elastic reflection waveform inversion

The 2D isotropic elastic-wave equation in frequency-domain (Aki and Richards, 2002) can be expressed as

$$-\omega^2\rho u_x = \frac{\partial}{\partial x}\left((\lambda+2\mu)\frac{\partial u_x}{\partial x}+\lambda\frac{\partial u_z}{\partial z}\right)+\frac{\partial}{\partial z}\left(\mu\frac{\partial u_z}{\partial x}+\mu\frac{\partial u_x}{\partial z}\right)+f_x \tag{1a}$$

$$-\omega^2\rho u_z = \frac{\partial}{\partial z}\left(\lambda\frac{\partial u_x}{\partial x}+(\lambda+2\mu)\frac{\partial u_z}{\partial z}\right)+\frac{\partial}{\partial x}\left(\mu\frac{\partial u_z}{\partial x}+\mu\frac{\partial u_x}{\partial z}\right)+f_z \tag{1b}$$

in which ω denotes the angular frequency, λ, μ indicate Lamé parameters and ρ is the density. u_x, u_z indicate the horizontal and vertical components of the displacement wavefield.

We can reformulate Eq. (1) in a brief formulation as

$$\mathbf{F}(\mathbf{m})\mathbf{u} = \mathbf{f}, \tag{2}$$

where

$$\mathbf{F}(\mathbf{m}) = \begin{bmatrix} -\omega^2\rho - \frac{\partial}{\partial x}\left((\lambda+2\mu)\frac{\partial}{\partial x}\right) - \frac{\partial}{\partial z}\left(\mu\frac{\partial}{\partial z}\right) & -\frac{\partial}{\partial x}\left(\lambda\frac{\partial}{\partial z}\right) - \frac{\partial}{\partial z}\left(\mu\frac{\partial}{\partial x}\right) \\ -\frac{\partial}{\partial x}\left(\mu\frac{\partial}{\partial z}\right) - \frac{\partial}{\partial z}\left(\lambda\frac{\partial}{\partial x}\right) & -\omega^2\rho - \frac{\partial}{\partial x}\left(\mu\frac{\partial}{\partial x}\right) - \frac{\partial}{\partial z}\left((\lambda+2\mu)\frac{\partial}{\partial z}\right) \end{bmatrix}, \tag{3}$$

$\mathbf{m} = (\lambda,\mu)^{\mathrm{T}}$, and $\mathbf{u} = (u_x, u_z)^{\mathrm{T}}$ represent the displacement wavefield. To simplify the problem, a constant density is assumed in this study.

Under the perturbation theory, the model parameters and wavefileds are composed of background part and perturbed part as $\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m}$, $\mathbf{u} = \mathbf{u}_0 + \delta\mathbf{u}$. Specifically, in terms of the Lamé parameters and wavefield components, the expressions can be written as

$$\lambda = \lambda_0 + \delta\lambda, \mu = \mu_0 + \delta\mu, u_x = u_{x0} + \delta u_x, u_z = u_{z0} + \delta u_z \tag{4}$$

Thus, the perturbed wavefield equation is obtained by subtracting the background wavefield from the total wavefield and applying the Born approximation:

$$\mathbf{F}(\mathbf{m}_0)\delta\mathbf{u} = \mathbf{B}(\delta\mathbf{m})\mathbf{u}_0 \tag{5}$$

where

$$\mathbf{B}(\delta\mathbf{m}) = \mathbf{A}_1\delta\lambda\mathbf{A}_\lambda + \mathbf{A}_1\delta\mu\mathbf{A}_\mu + \mathbf{A}_2\delta\mu\mathbf{A}'_\mu, \tag{6}$$

and

$$\mathbf{A}_1 = \begin{bmatrix} \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial z} \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} \frac{\partial}{\partial z} & 0 \\ 0 & \frac{\partial}{\partial x} \end{bmatrix},$$
$$\mathbf{A}_\lambda = \begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial z} \end{bmatrix}, \quad \mathbf{A}_\mu = \begin{bmatrix} 2\frac{\partial}{\partial x} & 0 \\ 0 & 2\frac{\partial}{\partial z} \end{bmatrix}, \quad \mathbf{A}'_\mu = \begin{bmatrix} \frac{\partial}{\partial z} & \frac{\partial}{\partial x} \\ \frac{\partial}{\partial z} & \frac{\partial}{\partial x} \end{bmatrix}. \tag{7}$$

A joint ERWI objective function (Wu and Alkhalifah, 2015; Li et al., 2019) is employed, integrating both incident and reflected waves:

$$J(\mathbf{m}_0, \delta\mathbf{m}) = \frac{1}{2}\|\mathbf{u}_0 + \delta\mathbf{u} - \mathbf{u}_{\mathrm{obs}}\|_2^2 \tag{8}$$

Incident waves and reflected waves correspond to the background and perturbed wavefields, respectively. The joint ERWI objective function leverages incident waves to improve the reconstruction of shallow velocities, thereby further enhancing the inversion of deeper velocity models. The Fréchet derivatives quantify the sensitivity of the seismic data to variations in the model parameters. Reflection waveform inversion adopts a scale separation strategy between the background and perturbed models. For the objective function defined in Eq. (8), it is essential to analyze the contributions of incident and reflected waves to the background and perturbation models. A detailed derivation of the Fréchet derivatives is provided in Appendix A.

The sensitivity kernel tensor corresponding to the perturbed model parameters is expressed as

$$\mathbf{L}_{\delta\mathbf{m}} = \begin{bmatrix} \mathbf{L}_{\delta\lambda} & \mathbf{L}_{\delta\mu} \end{bmatrix}. \tag{9}$$

The Fréchet derivatives with respect to the perturbed model parameters are given by

$$\mathbf{L}_{\delta\lambda} = \mathbf{G}_0\,\mathbf{A}_1\,\mathbf{A}_\lambda\,\mathbf{u}_0, \tag{10a}$$

$$\mathbf{L}_{\delta\mu} = \mathbf{G}_0\left(\mathbf{A}_1\mathbf{A}_\mu + \mathbf{A}_2\mathbf{A}'_\mu\right)\mathbf{u}_0. \tag{10b}$$

The sensitivity kernel tensor corresponding to the background model parameters is expressed as

$$\mathbf{L}_{m_0} = \begin{bmatrix} \mathbf{L}_{\lambda_0} & \mathbf{L}_{\mu_0} \end{bmatrix}. \tag{11}$$

The Fréchet derivatives with respect to the background model parameters are given by

$$\mathbf{L}_{\lambda_0} = \mathbf{G}_0\mathbf{A}_1\mathbf{A}_\lambda\mathbf{u}_0 + \mathbf{G}_0\mathbf{A}_1\mathbf{A}_\lambda\delta\mathbf{u} + \delta\mathbf{G}_\lambda\mathbf{A}_1\mathbf{A}_\lambda\mathbf{u}_0. \tag{12a}$$

$$\begin{aligned}\mathbf{L}_{\mu_0} &= \mathbf{G}_0\left(\mathbf{A}_1\mathbf{A}_\mu + \mathbf{A}_2\mathbf{A}'_\mu\right)\mathbf{u}_0 + \mathbf{G}_0\left(\mathbf{A}_1\mathbf{A}_\mu + \mathbf{A}_2\mathbf{A}'_\mu\right)\delta\mathbf{u} \\ &+ \delta\mathbf{G}_\mu\left(\mathbf{A}_1\mathbf{A}_\mu + \mathbf{A}_2\mathbf{A}'_\mu\right)\mathbf{u}_0.\end{aligned} \tag{12b}$$

Based on the adjoint-state method (derivation presented in Appendix B), the gradients of the objective function with respect to the perturbed model parameters are given by

$$\mathbf{g}(\delta\lambda) = -\sum_s \int \left[\left(\frac{\partial u_{x0}}{\partial x} + \frac{\partial u_{z0}}{\partial z}\right)\left(\frac{\partial \phi_x}{\partial x} + \frac{\partial \phi_z}{\partial z}\right)\right]\mathrm{d}\omega, \tag{13a}$$

$$\begin{aligned}\mathbf{g}(\delta\mu) &= -\sum_s \int \left[2\left(\frac{\partial u_{x0}}{\partial x}\frac{\partial \phi_x}{\partial x} + \frac{\partial u_{z0}}{\partial z}\frac{\partial \phi_z}{\partial z}\right)\right. \\ &\left. + \left(\frac{\partial u_{z0}}{\partial x} + \frac{\partial u_{x0}}{\partial z}\right)\left(\frac{\partial \phi_x}{\partial z} + \frac{\partial \phi_z}{\partial x}\right)\right]\mathrm{d}\omega.\end{aligned} \tag{13b}$$

For the background model parameters, the corresponding gradients can be expressed as

$$\begin{aligned}\mathbf{g}(\lambda_0) &= -\sum_s \int \left[\left(\frac{\partial u_{x0}}{\partial x} + \frac{\partial u_{z0}}{\partial z}\right)\left(\frac{\partial \phi_x}{\partial x} + \frac{\partial \phi_z}{\partial z}\right)\right. \\ &+ \left(\frac{\partial \delta u_x}{\partial x} + \frac{\partial \delta u_z}{\partial z}\right)\left(\frac{\partial \phi_x}{\partial x} + \frac{\partial \phi_z}{\partial z}\right) \\ &\left. + \left(\frac{\partial u_{x0}}{\partial x} + \frac{\partial u_{z0}}{\partial z}\right)\left(\frac{\partial \delta\phi_x}{\partial x} + \frac{\partial \delta\phi_z}{\partial z}\right)\right]\mathrm{d}\omega,\end{aligned} \tag{14a}$$

$$\begin{aligned}\mathbf{g}(\mu_0) = -\sum_s \int &\left[2\left(\frac{\partial u_{x0}}{\partial x}\frac{\partial \phi_x}{\partial x} + \frac{\partial u_{z0}}{\partial z}\frac{\partial \phi_z}{\partial z}\right)\right. \\ &+ \left(\frac{\partial u_{z0}}{\partial x} + \frac{\partial u_{x0}}{\partial z}\right)\left(\frac{\partial \phi_x}{\partial z} + \frac{\partial \phi_z}{\partial x}\right) \\ &+ 2\left(\frac{\partial \delta u_x}{\partial x}\frac{\partial \phi_x}{\partial x} + \frac{\partial \delta u_z}{\partial z}\frac{\partial \phi_z}{\partial z}\right) \\ &+ \left(\frac{\partial \delta u_z}{\partial x} + \frac{\partial \delta u_x}{\partial z}\right)\left(\frac{\partial \phi_x}{\partial z} + \frac{\partial \phi_z}{\partial x}\right) \\ &+ 2\left(\frac{\partial u_{x0}}{\partial x}\frac{\partial \delta\phi_x}{\partial x} + \frac{\partial u_{z0}}{\partial z}\frac{\partial \delta\phi_z}{\partial z}\right) \\ &\left. + \left(\frac{\partial u_{z0}}{\partial x} + \frac{\partial u_{x0}}{\partial z}\right)\left(\frac{\partial \delta\phi_x}{\partial z} + \frac{\partial \delta\phi_z}{\partial x}\right)\right]\mathrm{d}\omega.\end{aligned} \tag{14b}$$

here, ϕ_x and ϕ_z denote the horizontal and vertical components of the displacement wavefield obtained by backpropagating the data residuals, referred to as the adjoint wavefield. Similarly, $\delta\phi_x$ and $\delta\phi_z$ denote the horizontal and vertical components of the perturbed displacement wavefield, obtained by applying Born modeling to the adjoint wavefield, referred to as the adjoint perturbed wavefield. These wavefields can be computed by solving Eq. (B.5).

In elastic wave equation inversion problems, velocity is typically used as the model parameter. The P- and S-wave velocities are defined as

$$v_P = \sqrt{\frac{(\lambda+\mu)}{\rho}},\ v_S = \sqrt{\frac{\mu}{\rho}}, \tag{15}$$

The Fréchet derivatives with respect to the velocity model can be obtained via the chain rule:

$$\mathbf{L}_{v_{p0}} = 2\rho_0 v_{p0}\mathbf{L}_{\lambda_0},\quad \mathbf{L}_{v_{s0}} = -4\rho_0 v_{s0}\mathbf{L}_{\lambda_0} + 2\rho_0 v_{s0}\mathbf{L}_{\mu_0}, \tag{16a}$$

$$\mathbf{L}_{\delta v_p} = 2\rho_0 v_{p0}\mathbf{L}_{\delta\lambda},\quad \mathbf{L}_{\delta v_s} = -4\rho_0 v_{s0}\mathbf{L}_{\delta\lambda} + 2\rho_0 v_{s0}\mathbf{L}_{\delta\mu}. \tag{16b}$$

Similarly, the gradients with respect to the velocity model are expressed as

$$\mathbf{g}(v_{p0}) = 2\rho_0 v_{p0}\mathbf{g}(\lambda_0),\quad \mathbf{g}(v_{s0}) = -4\rho_0 v_{s0}\mathbf{g}(\lambda_0) + 2\rho_0 v_{s0}\mathbf{g}(\mu_0), \tag{17a}$$

$$\mathbf{g}(\delta v_p) = 2\rho_0 v_{p0}\mathbf{g}(\delta\lambda),\quad \mathbf{g}(\delta v_s) = -4\rho_0 v_{s0}\mathbf{g}(\delta\lambda) + 2\rho_0 v_{s0}\mathbf{g}(\delta\mu). \tag{17b}$$

By applying these gradient expressions in combination with a gradient-based optimization algorithm, ERWI can be performed to invert for the velocity models. In ERWI, updating the background velocity models requires information from the perturbation models. Consequently, it is common practice either to update the perturbation and background models alternately, or to construct a fixed perturbation model first and then update the background models based on it.

2.2. Improvement scheme based on Hessian and scale fusion strategy

In the ERWI process, the construction of the perturbation models is crucial, as they not only contribute to the overall velocity models update but also influence the update of the background models. The gradients of the background models receive contributions from both transmitted and reflected waves, with the reflected waves arising from the scattering of the background wavefields within the perturbation models. Therefore, enhancing the resolution and accuracy of the perturbation models is essential for improving the final ERWI results.

For multi-parameter ERWI, the Hessian operator plays a key role in the optimization process by incorporating second-order derivative information of the objective function. It characterizes the relationship between changes in model parameters and the resulting data misfit, describing the curvature of the objective function in the model parameter space. This allows for more informed and balanced updates during inversion. Moreover, the multi-parameter Hessian accounts for the differing sensitivities of model parameters, balancing updates and mitigating scaling issues that may arise from heterogeneity in model properties.

The Hessian matrix can be obtained by taking the second-order partial derivatives of the objective function in Eq. (8). With the model scale separation in ERWI, the Hessian is divided into components corresponding to the background model parameters and the perturbation model parameters. Given the significance of the perturbation models and the associated computational cost, we focus on the Hessian calculation with respect to the perturbation models. The Hessian matrix corresponding to the perturbation models can be expressed as:

$$\mathbf{H}_{\delta m} = \frac{\partial^2 J(\mathbf{m}_0, \delta\mathbf{m})}{\partial\delta\mathbf{m}^2} = \mathbf{L}_{\delta m}^{\mathrm{T}}\mathbf{L}_{\delta m} + \left(\frac{\partial\mathbf{L}_{\delta m}}{\partial\delta\mathbf{m}}\right)^{\mathrm{T}}\Delta\mathbf{u}. \tag{18}$$

By neglecting the nonlinear terms, the Gauss-Newton approximate Hessian is given by

$$\mathbf{H}_{\delta m}^{a} = \mathbf{L}_{\delta m}^{\mathrm{T}}\mathbf{L}_{\delta m}. \tag{19}$$

Expanding the objective function in Eq. (8) around $\delta\mathbf{m}_0$ using a second-order Taylor series yields

$$J(\mathbf{m}_0, \delta\mathbf{m}) \approx J(\mathbf{m}_0, \delta\mathbf{m}_0) + \frac{\partial J(\mathbf{m}_0, \delta\mathbf{m})}{\partial\delta\mathbf{m}}\Delta\delta\mathbf{m} + \frac{1}{2!}\frac{\partial^2 J(\mathbf{m}_0, \delta\mathbf{m})}{\partial(\delta\mathbf{m})^2}(\Delta\delta\mathbf{m})^2 \tag{20}$$

Substituting Eq. (19) into Eq. (20) and setting the derivative with respect to $\Delta\delta\mathbf{m}$ to zero, the Gauss-Newton update direction satisfies

$$\mathbf{H}_{\delta m}^{a}\Delta\delta\mathbf{m} = -\mathbf{g}(\delta\mathbf{m}) \tag{21}$$

The linear system in Eq. (21) can be solved using a matrix-free conjugate gradient algorithm (Saad, 2003), where the critical step is the computation of the Hessian-vector product. To further alleviate the effects of uneven illumination, a diagonal preconditioning operator is introduced and applied to Eq. (21):

$$\mathbf{W}\mathbf{H}_{\delta m}^{a}\Delta\delta\mathbf{m} = -\mathbf{W}\mathbf{g}(\delta\mathbf{m}) \tag{22}$$

where $\mathbf{W}$ is the illumination operator constructed from the source-side wavefield autocorrelation.

The Hessian-vector product under the Gauss-Newton approximation can be expressed as

$$\mathbf{H}_{\delta m}^{a}\mathbf{p} = \mathbf{L}_{\delta m}^{\mathrm{T}}\mathbf{L}_{\delta m}\mathbf{p}. \tag{23}$$

where $\mathbf{p}$ is a vector in the model space, typically initialized as the current model update direction in the matrix-free conjugate gradient algorithm. The detailed computation of the Hessian-vector product is provided in Appendix C. With this efficient computation, the Gauss-Newton update direction can be obtained using the matrix-free conjugate gradient algorithm as shown in Algorithm (1).

Algorithm 1. Matrix-free conjugate gradient

Input: $\mathbf{g}(\delta\mathbf{m})$, $t = 0$, $\Delta\delta\mathbf{m} = 0$, $\mathbf{r}_0 = \mathbf{W}\mathbf{H}_{\delta m}^{a}\Delta\delta\mathbf{m} + \mathbf{W}\mathbf{g}(\delta\mathbf{m})$, $\mathbf{p}_0 = -\mathbf{r}_0$
while stopping criterion is not satisfied **do**

$\alpha_t = \dfrac{\mathbf{r}_t^{\mathrm{T}}\mathbf{r}_t}{\mathbf{p}_t^{\mathrm{T}}\mathbf{W}\mathbf{H}_{\delta m}^{a}\mathbf{p}_t}$

$\Delta\delta\mathbf{m}^{t+1} = \Delta\delta\mathbf{m}^{t} + \alpha_t\mathbf{p}_t$

$\mathbf{r}_{t+1} = \mathbf{r}_t + \alpha_t\mathbf{W}\mathbf{H}_{\delta m}^{a}\mathbf{p}_t$

$\beta_{t+1} = \dfrac{\mathbf{r}_{t+1}^{\mathrm{T}}\mathbf{r}_{t+1}}{\mathbf{r}_t^{\mathrm{T}}\mathbf{r}_t}$

$\mathbf{p}_{t+1} = -\mathbf{r}_{t+1} + \beta_{t+1}\mathbf{p}_t$

$t = t + 1$

end while
Output: $\Delta\delta\mathbf{m}$

The truncated Gauss-Newton (TGN) process consists of two nested loops: the inner loop computes the Gauss-Newton direction using the matrix-free conjugate gradient method, while the outer loop updates the model and computes the gradients. The step size for model updates is determined using a parabolic fitting method (Vigh and Starr, 2008).

The stopping criteria of inner loops is given by

$$\|\mathbf{W}\mathbf{H}^{a}_{\delta m}\Delta\delta\mathbf{m}+\mathbf{W}\mathbf{g}(\delta\mathbf{m})\|_2 < \epsilon\|\mathbf{W}\mathbf{g}(\delta\mathbf{m})\|_2, \tag{24}$$

where ϵ controls accuracy in the search direction (Eisenstat and Walker, 1996). To ensure computational efficiency and prevent overfitting of the linear system, a maximum of 5 inner iterations is imposed in addition to the stopping criterion.

Fig. 1 illustrates the commonly used alternating update workflow in reflection waveform inversion. In this workflow, the model is separated into background and perturbation components based on scale separation, and each component is updated individually. After a certain number of iterations, the two components are combined to produce the final inversion results. An alternative approach constructs a temporary perturbation model at each iteration, and after a specified number of iterations, the updated background model is taken as the final inversion result.

The first strategy, which alternately updates the perturbation and background models after scale separation, can fail when the constructed perturbation models are insufficiently accurate or when the inversion becomes trapped in local minima. Furthermore, differences in the wavenumber ranges of the background and perturbation models may introduce numerical fluctuations in the final results when the two models are summed. The second strategy, which constructs temporary perturbation models, discards portions of the perturbation information that could otherwise facilitate the inversion update, resulting in inversion results that lack high-wavenumber details and exhibit slower convergence.

Based on this analysis, we propose the workflow illustrated in Fig. 2, which employs a scale fusion strategy combined with TGN optimization method to overcome the limitations of the aforementioned workflows.

Based on the derivation in Section 2.1, it can be observed that the Fréchet derivatives of the background and perturbation models share common terms. The first term of the Fréchet derivative for the background models corresponds to the Fréchet

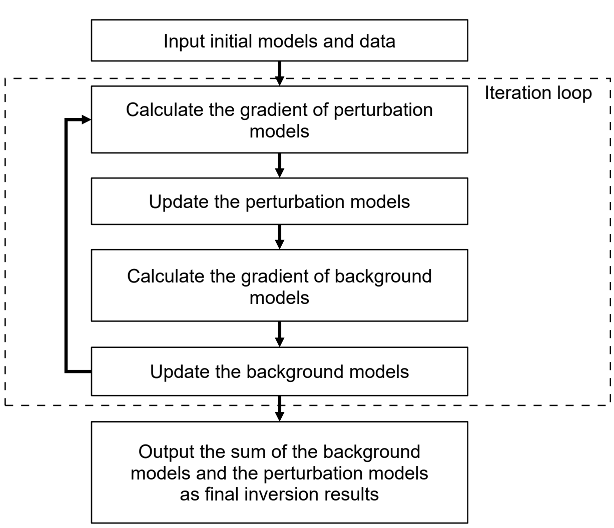

Fig. 1. The workflow of conventional ERWI with alternating updates.

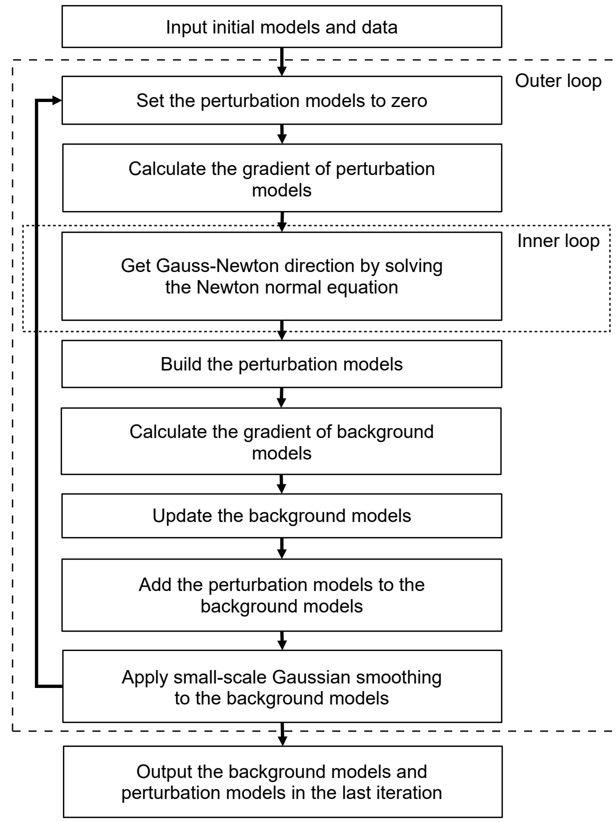

Fig. 2. The workflow of Hessian-based ERWI with scale fusion strategy.

derivative for the perturbation models. This observation motivates an integration within the ERWI workflow to stabilize the inversion update process and improve convergence speed.

In the proposed strategy, the TGN method is used to construct temporary perturbation models. These perturbation models are then employed to compute the gradients of the background models, which are subsequently updated. The perturbation models are then added back to the background models. The similarity between the Fréchet derivatives of the background and perturbation models provides a theoretical justification for this approach. To ensure stable updates of the background models in the low-to mid-wavenumber ranges, a small-scale Gaussian smoothing is applied, while the high-wavenumber information is derived solely from the perturbation models. The iteration proceeds to the next cycle with the perturbation models reset to zero, and the process is repeated.

In conventional ERWI, the model is updated by alternately updating the perturbation and background models, where two iterations constitute one complete model update. The wavefield reconstruction technique is applied throughout the process. Under the employed objective function, computing the gradient of the perturbation model requires 4 wave equation simulations, the background model gradient requires 6 simulations, and the line search step-size estimation involves 6 simulations. Hence, one perturbation model update iteration requires 10 simulations, while one background model update iteration requires 12 simulations, resulting in a total of 22 simulations for a complete update (two iterations).

In the proposed ERWI method, the perturbation model is updated using the Gauss-Newton approach. Each inner loop iteration for computing the Hessian-vector product requires 4 simulations, and assuming a maximum of 5 inner iterations, the inner loop consumes 20 simulations. Including 10 simulations for additional gradient and step-size computations, the outer iteration for updating the

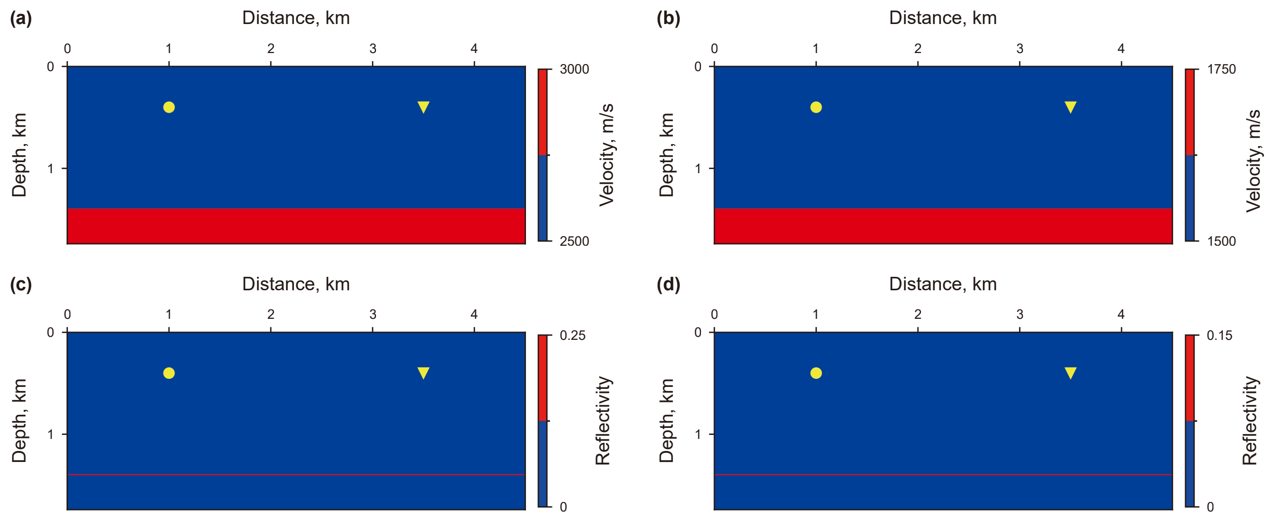

Fig. 3. The two-layered model. **(a)** shows P-wave velocity, **(b)** shows S-wave velocity, **(c)** shows P-wave reflectivity, and **(d)** shows S-wave reflectivity. The circle represents the source location, and the inverted triangle represents the receiver location, forming a source-receiver pair.

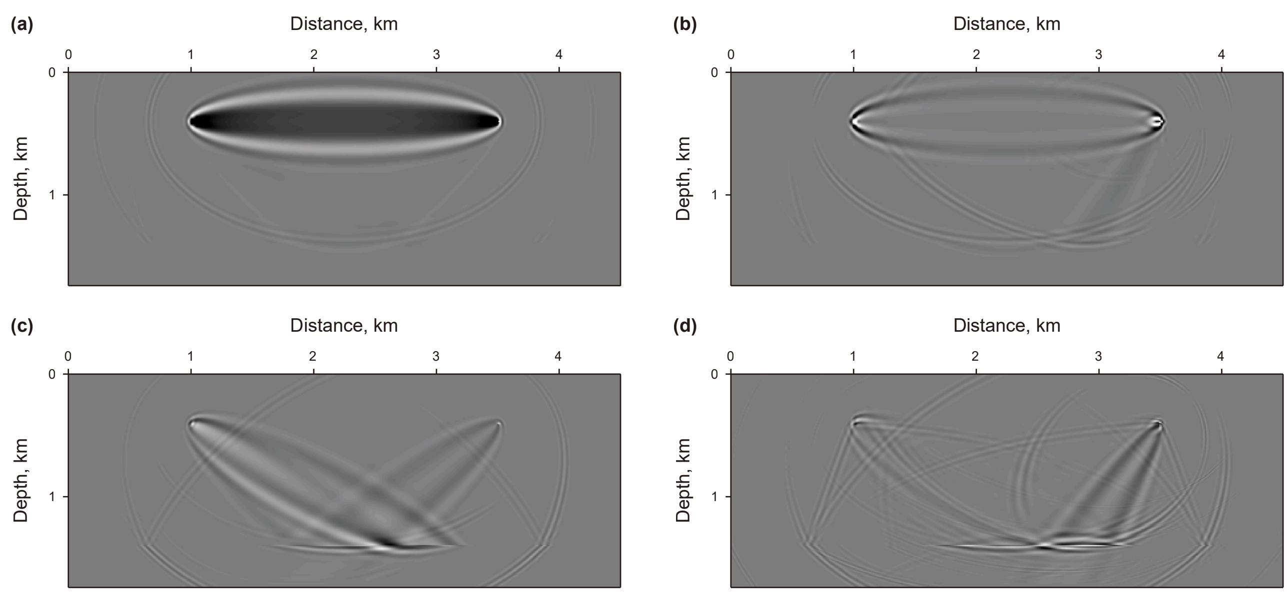

Fig. 4. Sensitivity kernels for a single source-receiver pair in the two-layer model. **(a)** and **(b)** show the P-wave and S-wave sensitivity kernels in EFWI, which also correspond to the first terms of the P-wave and S-wave background model sensitivity kernels in ERWI. **(c)** and **(d)** show the last two terms of the P-wave and S-wave background model sensitivity kernels in ERWI, representing the reflection wave sensitivity kernel components.

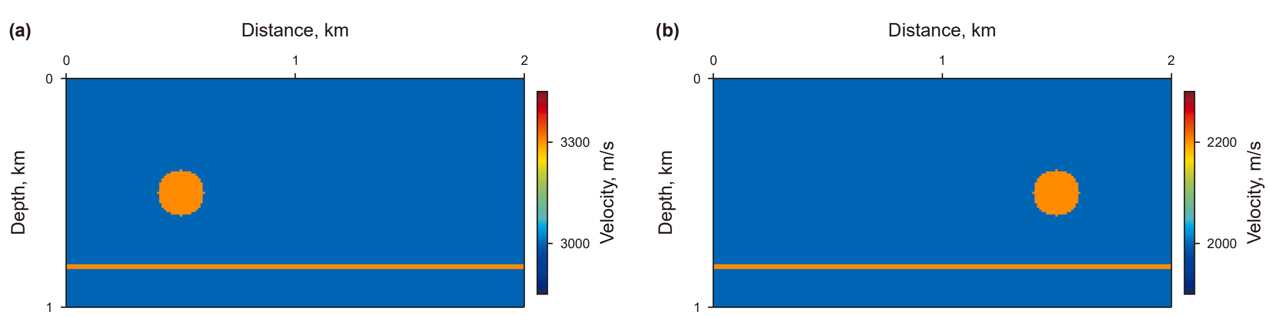

Fig. 5. The true velocity model of the anomaly model, where **(a)** is the P-wave velocity and **(b)** is the S-wave velocity.

perturbation model requires 30 simulations in total. The background model update remains unchanged from the conventional method, requiring 12 simulations. Consequently, one full update (two iterations) in the proposed method involves up to 42 wave equation simulations. Therefore, the computational cost of the proposed ERWI approach is less than twice that of the conventional ERWI method.

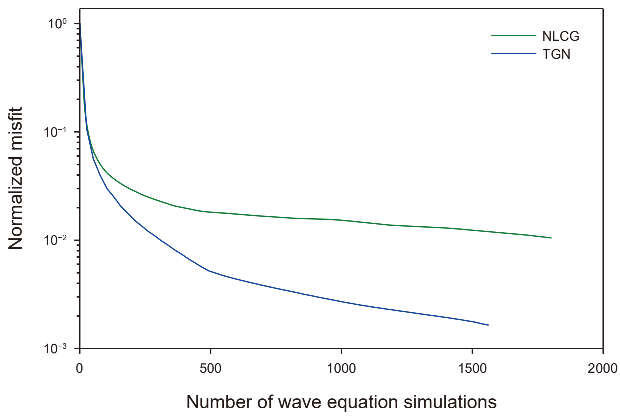

Fig. 6. The convergence curves in the anomaly model.

3. Numeriacal experiments

3.1. Fréchet derivative analysis

The Fréchet derivative plays a crucial role in the optimization process of ERWI, serving as a fundamental component for model parameter updates. It quantifies the sensitivity of the observed data to small perturbations in the model parameters, revealing the relationship between the model and the data. Specifically, the Fréchet derivative characterizes how a change in the model affects the predicted data, and provides the basis for calculating the gradient.

During the inversion process, the Fréchet derivative not only underpins optimization methods such as gradient descent but also offers insight into how different model parameters influence data fitting. In complex media or nonlinear problems, the Fréchet derivative guides the inversion algorithm, helping to avoid local minima and inappropriate update directions, thereby enhancing the accuracy and efficiency of the inversion.

Thus, the Fréchet derivative is of fundamental importance in ERWI. It forms the basis for gradient computation and model updates while providing theoretical support for parameter selection

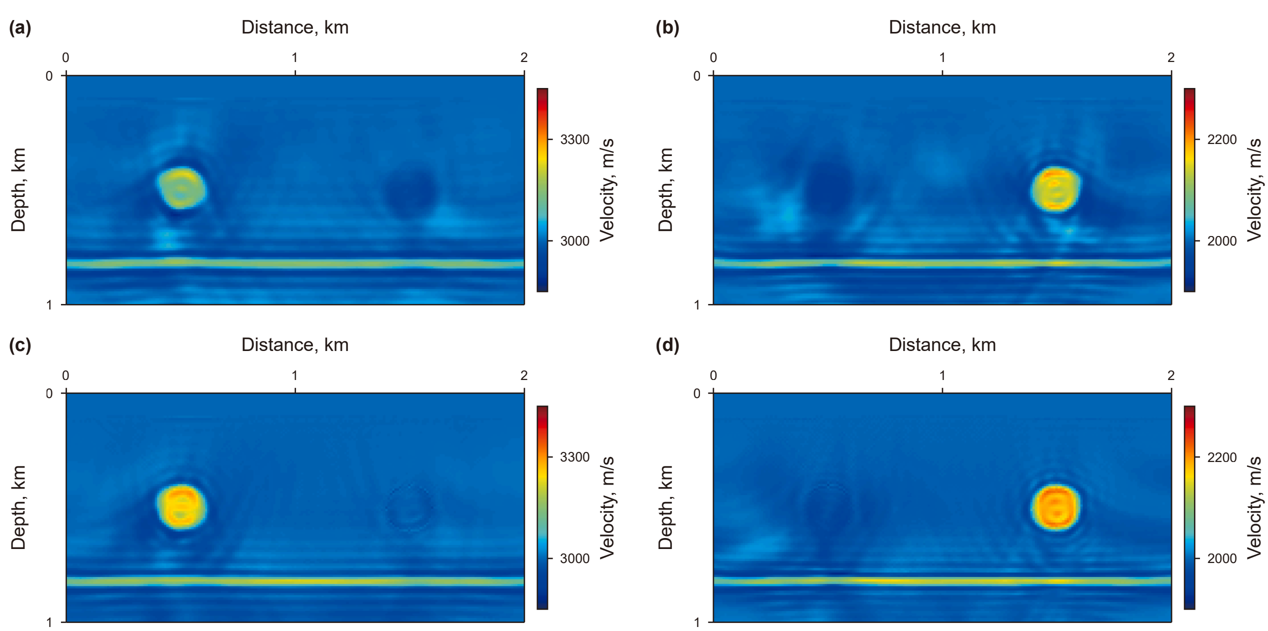

Fig. 7. The inversion results of the anomaly model. **(a)** and **(b)** represent the P-wave and S-wave velocities based on the NLCG method, while **(c)** and **(d)** represent the P-wave and S-wave velocities based on the TGN method.

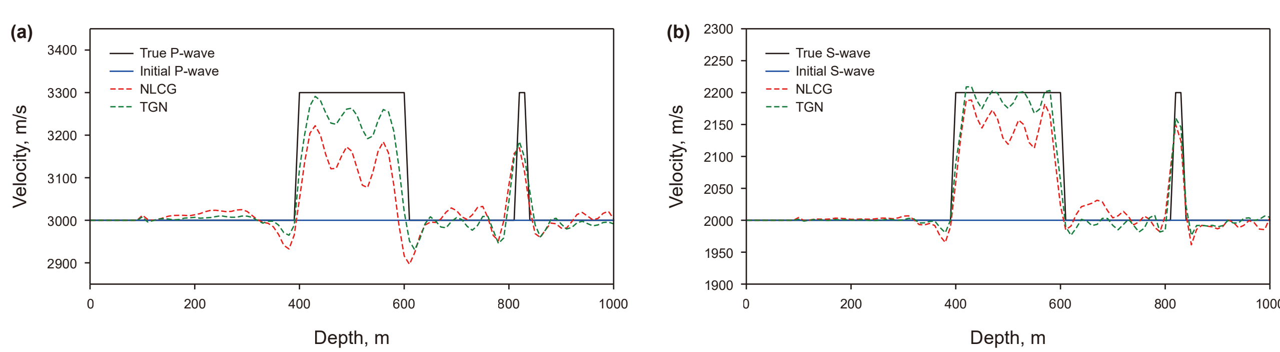

Fig. 8. The vertical velocity profiles extracted through the anomaly centers: **(a)** the P-wave velocity profile at a horizontal distance of 0.5 km, and **(b)** the S-wave velocity profile at a horizontal distance of 1.5 km.

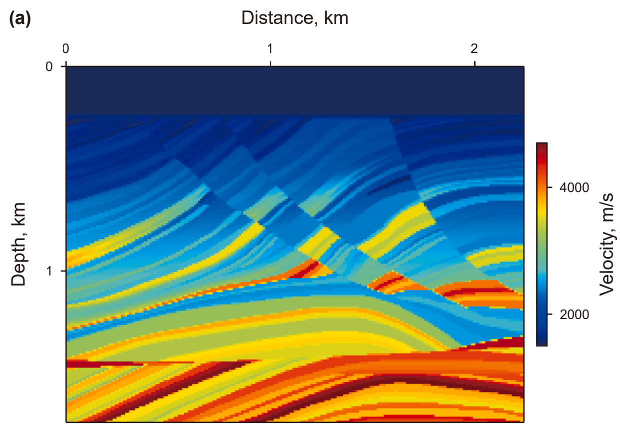

Fig. 9. The true velocity model of the Marmousi2 model, where **(a)** is the P-wave velocity and **(b)** is the S-wave velocity.

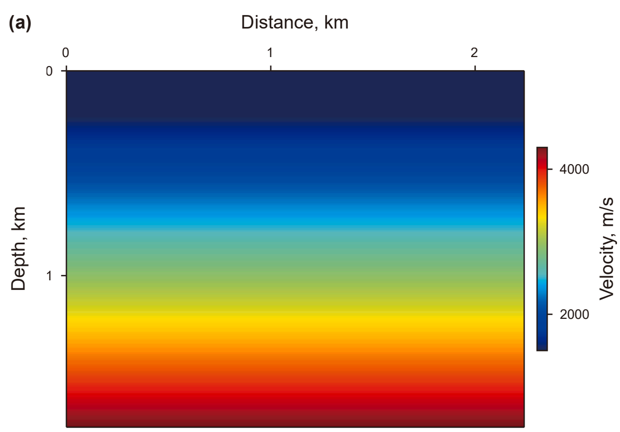

Fig. 10. The initial velocity model of the Marmousi2 model, where **(a)** is the P-wave velocity and **(b)** is the S-wave velocity.

during optimization. By properly exploiting the Fréchet derivative, ERWI can effectively recover accurate subsurface models from seismic data.

The two-layer model illustrated in Fig. 3 is used to demonstrate the effect of the Fréchet derivative, i.e., the sensitivity kernel, on the inversion update process in ERWI. Fig. 4 shows the sensitivity kernels for different components in ERWI. Fig. 4(a) shows the P-wave sensitivity kernel in EFWI, which also corresponds to the first term of the P-wave background model sensitivity kernel in ERWI and to the perturbation model sensitivity kernel in ERWI. Similarly, Fig. 4(b) depicts the S-wave sensitivity kernel in EFWI. From Fig. 4(a) and (b), it is evident that the transmitted wavepaths dominate the energy in EFWI sensitivity kernels, particularly for the P-wave, where the tomographic contributions of reflected wavepaths are relatively weak. Fig. 4(c) and (d) show the P-wave and S-wave reflection wave sensitivity kernels in ERWI, demonstrating that the reflection wave sensitivity kernels effectively update the model parameters along reflection wavepaths using the tomographic components of reflected waves.

3.2. Hessian crosstalk reduction

We analyze multi-parameter crosstalk reduction using the Hessian-based method on an anomaly model. This experiment is designed to demonstrate the effectiveness of the Hessian in mitigating multi-parameter crosstalk effects, so background model updating is not required, and ERWI is therefore reduced to EFWI. Results obtained using the nonlinear conjugate gradient (NLCG) method are compared with those from the TGN method. The true velocity model is shown in Fig. 5, and the initial model consists of a homogeneous background velocity. The model dimensions are 201×101, with both horizontal and vertical grid spacings of 10 m. Shots are spaced 50 m apart, with a total of 41 shots, and receivers are spaced every 10 m, totaling 201 receivers. A 10 Hz Ricker wavelet is used as the source to generate multicomponent seismic data, which serve as the observed data.

In this experiment, each update of the NLCG method requires 6 wave equation simulations, whereas each update of the TGN method requires up to 26 simulations. Fig. 6 shows the convergence curves of the data misfit versus the number of wave equation simulations for the two methods, providing an intuitive comparison of their computational cost and data misfit reduction. The NLCG method is iterated 300 times, whereas the TGN method is iterated 60 times. Because the data misfit can decrease by several orders of magnitude for this simple model, the vertical axis is plotted on a logarithmic scale. It can be observed that the TGN method achieves a lower data misfit upon convergence. Fig. 7 compares the inversion results obtained after 300 iterations of the NLCG method and 60 iterations of the TGN method. The NLCG results for both P-wave and S-wave velocities exhibit strong crosstalk artifacts, whereas the TGN results show significantly reduced crosstalk between the two anomaly bodies. Moreover, the TGN inversion results contain fewer artifacts and noise, reflecting the advantage of explicitly accounting for Hessian effects. Further comparison is performed by extracting vertical single-trace profiles through the centers of the two anomalies. Fig. 8(a) shows the P-wave vertical profile at a horizontal position of 0.5 km, and Fig. 8(b) shows the S-wave profile at 1.5 km. It is evident that the TGN method better recovers the true anomaly velocities and produces fewer sidelobes on either side, indicating more accurate delineation of the anomaly boundaries.

3.3. Inversion of Marmousi2 model

The Marmousi2 model is used for numeriacal experiment to demonstrate the advantages of the proposed ERWI method. The true velocity model is shown in Fig. 9, while Fig. 10 presents the linearized initial velocity model adopted for ERWI, in which the velocity of the top layer is fixed. The model dimensions are 225 × 175, with a grid spacing of 10 m in both the horizontal and vertical directions. Sources and receivers are deployed along the model surface, with a shot spacing of 100 m, resulting in a total of 23 shots. The receiver spacing is 10 m, giving 225 receivers in total. Synthetic multicomponent seismic data are generated by forward modeling using a 10 Hz Ricker wavelet, and frequency components below 5 Hz are filtered out.

The inversion results obtained using EFWI are shown in Fig. 11. It is evident that, with a linear initial model and in the absence of long-offset data and low-frequency information, EFWI fails to yield

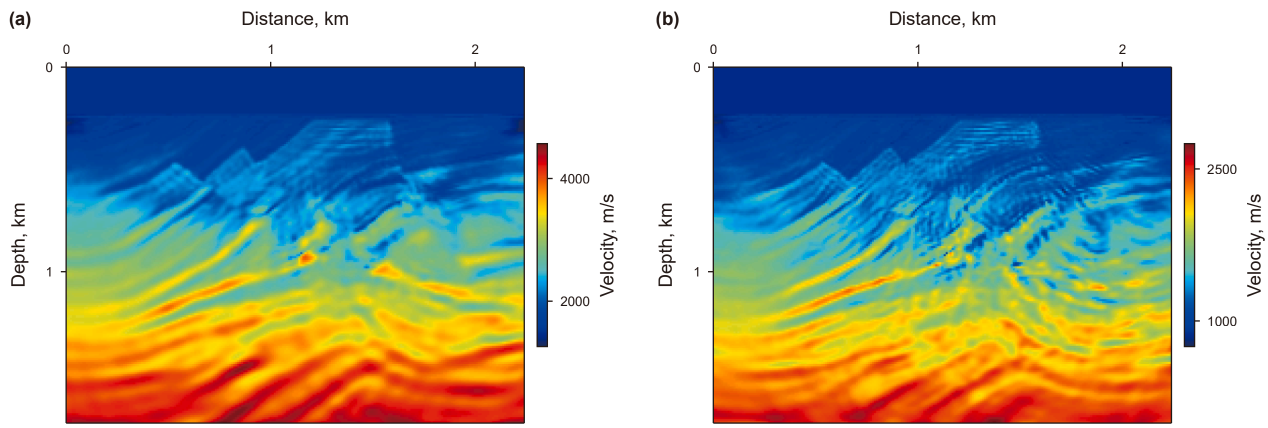

Fig. 11. The EFWI inversion results for the Marmousi2 model. Panels **(a)** and **(b)** show the P-wave and S-wave velocity models, respectively.

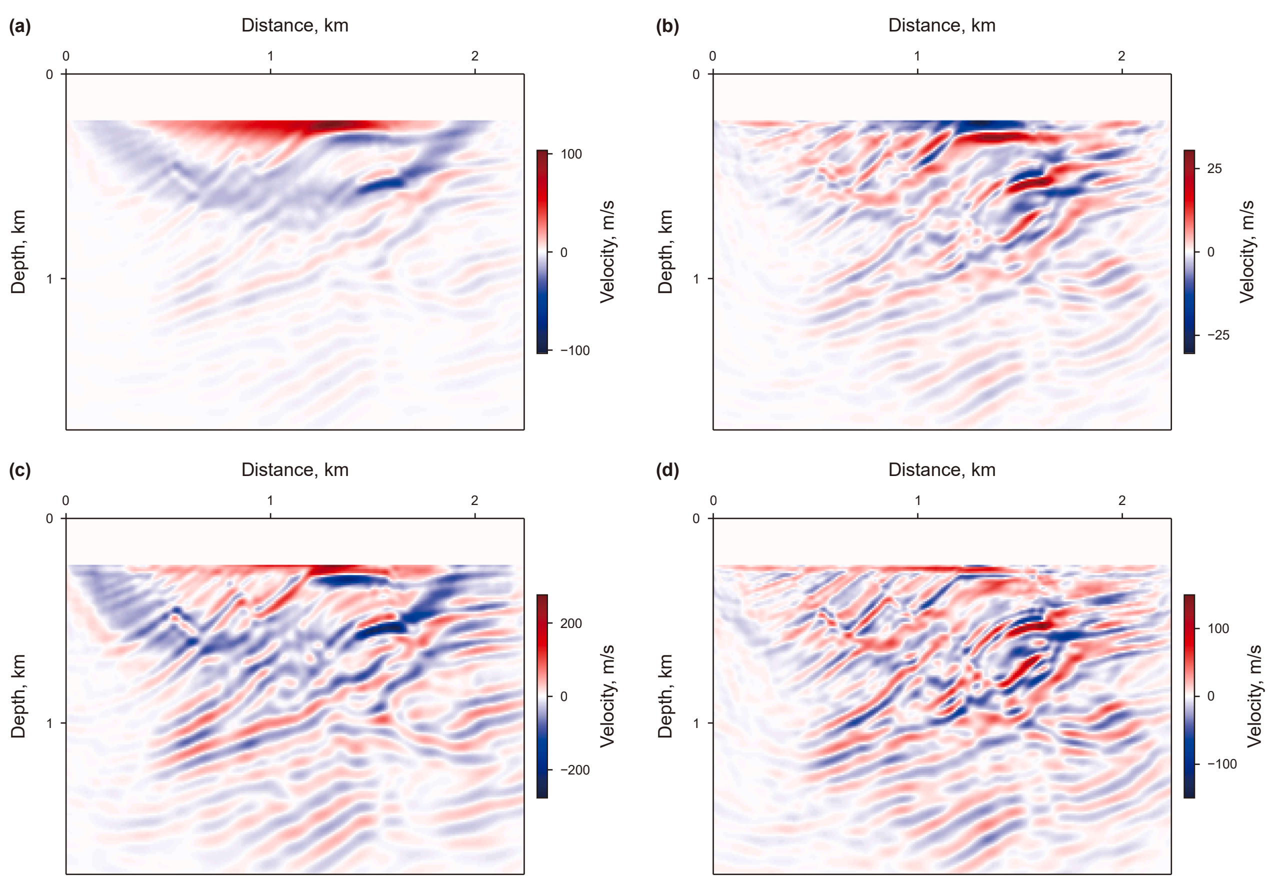

Fig. 12. The perturbation models at the first iteration. **(a)** and **(b)** represent the P-wave and S-wave velocity perturbation models obtained by the NLCG method, while **(c)** and **(d)** represent the P-wave and S-wave velocity perturbation models obtained by the TGN method.

satisfactory results. Under these conditions, ERWI can partially improve the performance of EFWI. The effectiveness of ERWI largely depends on the quality of the constructed perturbation model, which generates reflected-wave energy to update the background model. This observation motivates the construction of higher-quality perturbation models.

Fig. 12 presents the perturbation models generated in the first iteration using the NLCG and TGN methods. Owing to unbalanced illumination and limited bandwidth, the perturbation models produced by the NLCG method, as shown in Fig. 12(a) and (b), exhibit pronounced amplitude imbalance and noticeable artifacts. Furthermore, substantial amplitude discrepancies between the P-wave and S-wave perturbation models, arising from multi-parameter trade-off effects, result in inaccurate background model updates, slower convergence behavior, and a reduction in the final inversion accuracy. In contrast, the perturbation models generated by the TGN method (Fig. 12(c) and (d)) more clearly delineate the anomaly locations, compensate for illumination deficiencies induced by the acquisition geometry, and yield sharper perturbation models by alleviating the limited-bandwidth effects of the seismic data. In addition, by reducing multi-parameter trade-off effects, the TGN method achieves a better balance between the amplitudes of the P-wave and S-wave perturbation models. Collectively, these improvements facilitate faster convergence and lead to enhanced inversion accuracy.

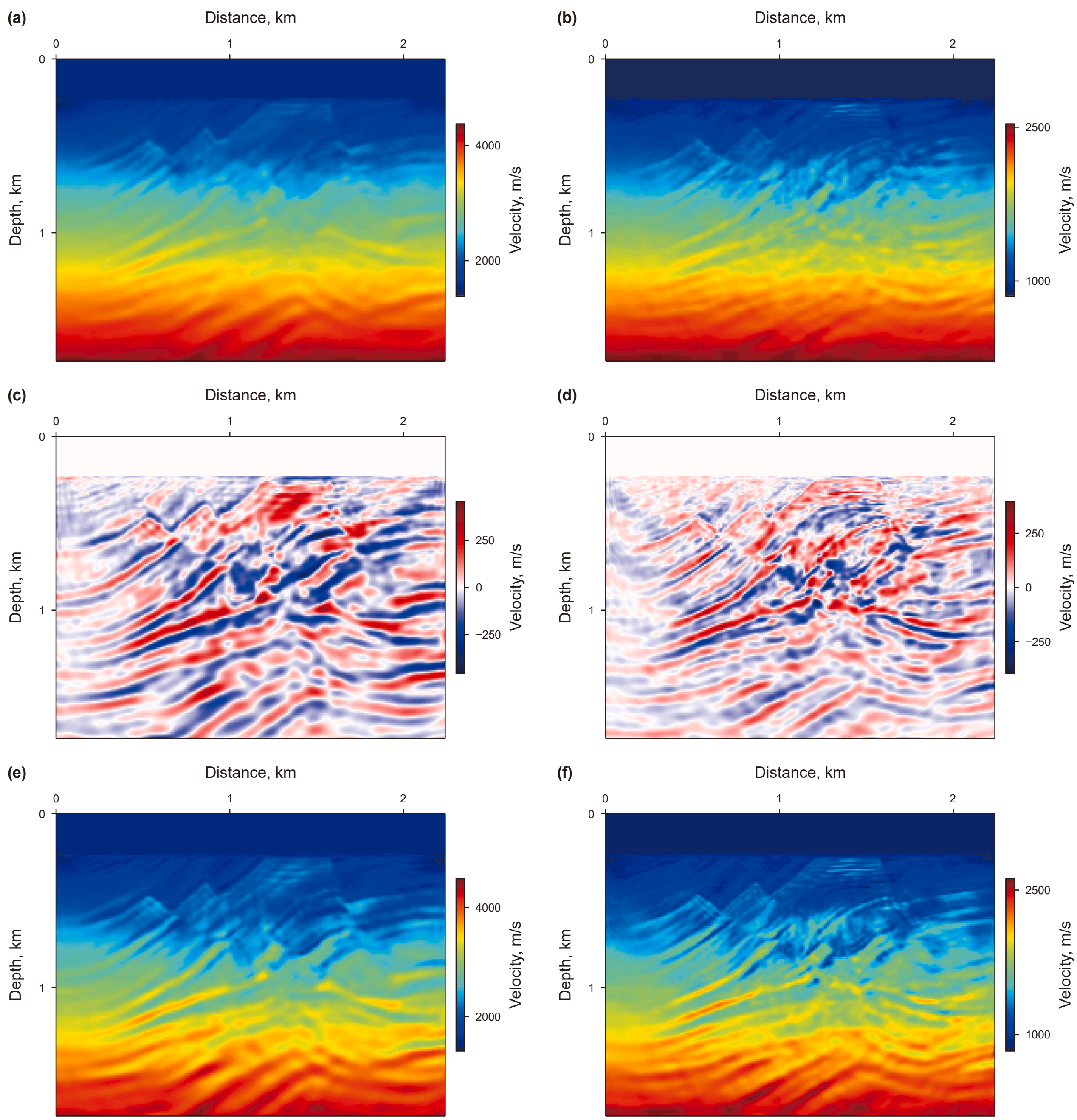

Fig. 13. The conventional ERWI inversion results for the Marmousi2 model. **(a)** and **(b)** show the background P-wave and S-wave velocity models, respectively; **(c)** and **(d)** show the corresponding perturbation models; **(e)** and **(f)** present the final inversion results, obtained by adding the background and perturbation models.

The conventional ERWI inversion results based on the NLCG method with the scale-separation strategy (referred to as method1 in the context) are shown in Fig. 13. As illustrated in Fig. 13(a) and (b), the inverted P-wave and S-wave background models are significantly improved relative to the linearized initial model, demonstrating the capability of conventional ERWI to construct a reasonable initial background model. Fig. 13(c) and (d) present the corresponding P-wave and S-wave perturbation models. The final inversion results obtained by summing the background and perturbation models are shown in Fig. 13(e) and (f). Overall, ERWI provides improved results compared with EFWI. However, the perturbation models still contain substantial artifacts, characterized by unbalanced amplitudes and illumination, poorly defined velocity interfaces, and excessively large update magnitudes. These deficiencies adversely affect the quality of the final summed models, as evident in Fig. 13(e) and (f).

The inversion results obtained using the proposed ERWI workflow based on the TGN method and the scale-fusion strategy (referred to as method2 in the context) are shown in Fig. 14. In this framework, the background model after inversion can be regarded as the final result, because the temporary perturbation models are explicitly incorporated into the background model at each iteration through scale fusion. For comparison, the perturbation models from the last iteration are also presented in Fig. 14(c) and (d). These perturbation models exhibit an enhanced ability to delineate anomalous interfaces, with more balanced amplitudes. Compared with the conventional ERWI method, the proposed method produces perturbation models with substantially higher resolution, clearer characterization of anomaly locations, and more effective suppression of multi-parameter crosstalk. The wavenumber spectra of the final perturbation models, shown in Fig. 15, further indicate that the proposed method recovers richer high-wavenumber content and a broader wavenumber bandwidth. A comparison between the final ERWI results from the conventional method (Fig. 13(e) and (f)) and those from the proposed method (Fig. 14(a) and (b)) shows that the latter achieves overall superior inversion quality, with fewer artifacts, better-preserved boundary updates, and more clearly defined interfaces, thereby leading to improved overall inversion performance.

Fig. 16 presents a comparison of the convergence curves for the three methods, where EFWI and conventional ERWI are iterated 250 times, and the proposed ERWI method is iterated 100 times. It can be seen that the proposed ERWI method converges to a lower data misfit with fewer wave equation simulations, demonstrating better convergence performance.

Fig. 17 presents the vertical profiles of the inversion results. Both EFWI and conventional ERWI exhibit noticeable numerical fluctuations and multiple instances of mispositioned structures. In contrast, the proposed ERWI method produces more stable results that are closer to the true values. Although the inversion results from the proposed ERWI method do not perfectly match the true velocities, they remain generally acceptable given the challenging conditions, including short-offset data, missing low-frequency components, and a linear initial model. Overall, the proposed ERWI method demonstrates a clear improvement over both EFWI and conventional ERWI.

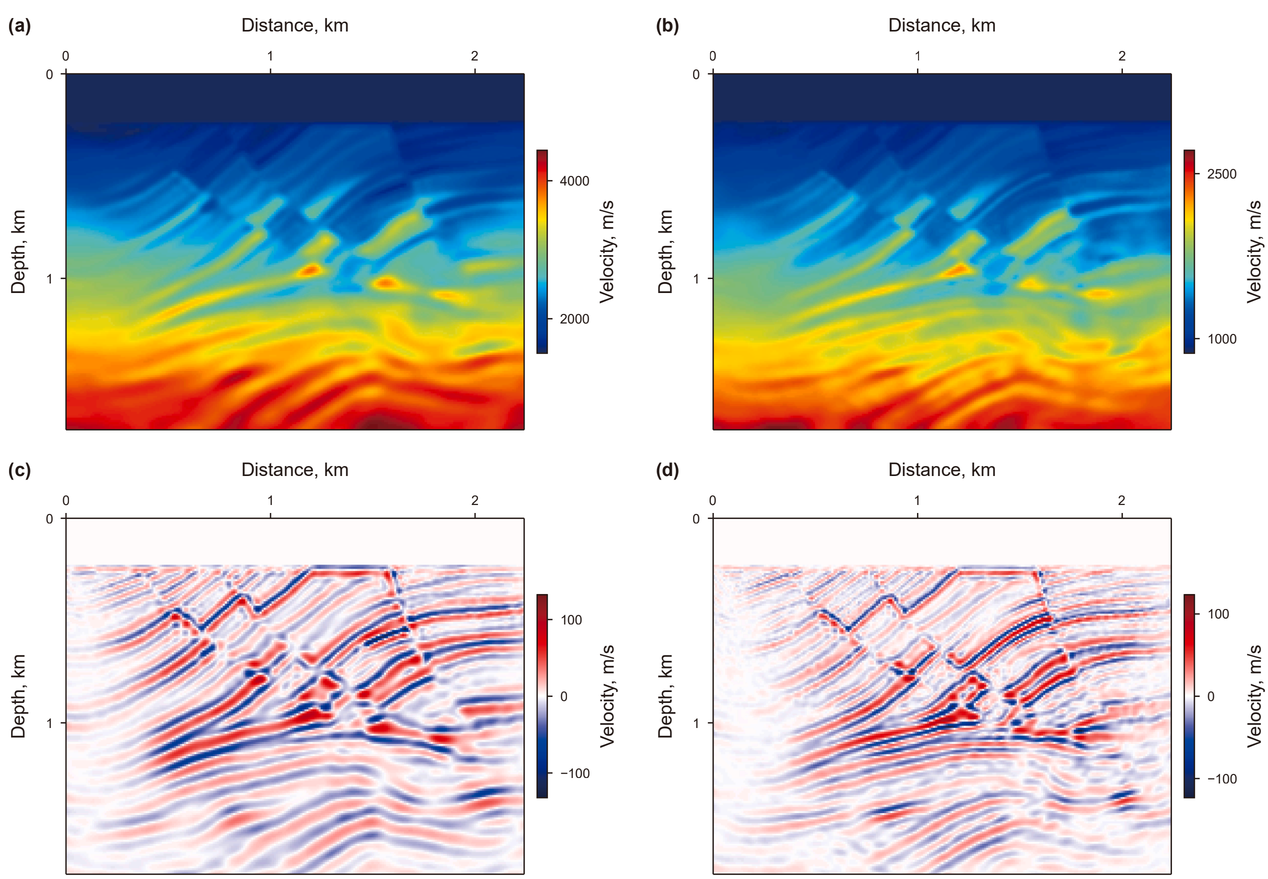

Fig. 14. The proposed ERWI inversion results of Marmousi2 model. **(a)** and **(b)** show the background P-wave and S-wave velocity models, which are also the final ERWI inversion results in the proposed method. **(c)** and **(d)** show the temporary perturbation P-wave and S-wave velocity models at the final iteration.

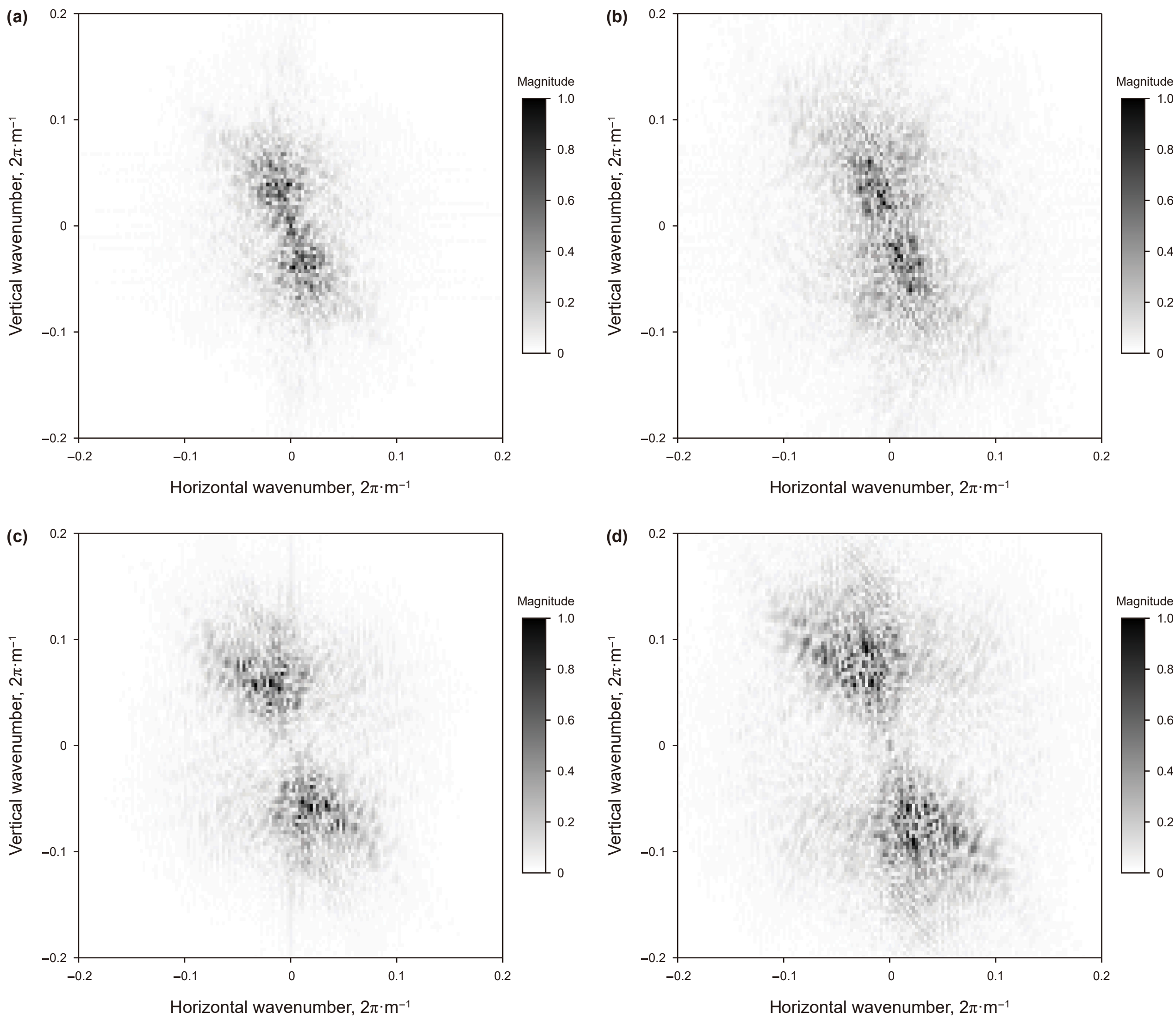

Fig. 15. The wavenumber spectra of the perturbation models at the final iteration. **(a)** and **(b)** show the P-wave and S-wave perturbation model spectra obtained using the conventional ERWI method, while **(c)** and **(d)** show the corresponding spectra from the proposed ERWI method.

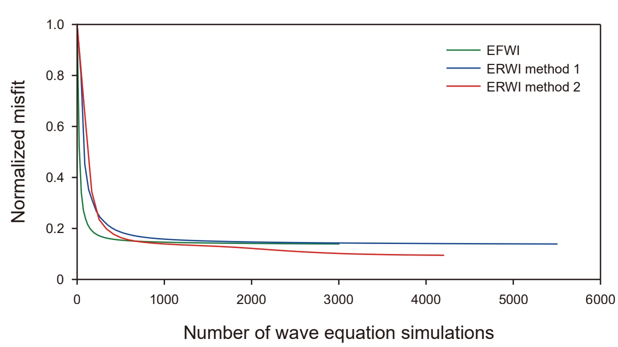

Fig. 16. The convergence curves in the Marmousi2 model.

To perform a quantitative analysis of the entire inversion results, the model error is defined as $\sqrt{\|\mathbf{m}_{\text{inv}} - \mathbf{m}_{\text{true}}\|_2^2 / \|\mathbf{m}_{\text{true}}\|_2^2}$. As shown in Table 1, the model error for the EFWI method increases relative to the initial model, indicating that the method becomes trapped in a local minimum under challenging conditions. In the conventional ERWI method, the P-wave velocity model error decreases compared to the initial model, whereas the S-wave velocity model error increases. This is partly due to parameter trade-off effects that hinder the accurate updating of the S-wave velocity model. In contrast, the proposed ERWI method substantially reduces both P-wave and S-wave velocity model errors. This improvement arises from updating the perturbation model using the Gauss-Newton direction, which incorporates the multi-parameter Hessian matrix to mitigate the coupling between P- and S-wave velocities. Moreover, the scale fusion strategy prevents error accumulation in the perturbation model, ensuring progressively more accurate updates that effectively guide the background model inversion. Consequently, the final inversion result obtained using the proposed ERWI method exhibits lower model error than both the EFWI and conventional ERWI methods, demonstrating its effectiveness in enhancing inversion accuracy.

3.4. Robustness to noise

To further evaluate the robustness of the proposed method, the ERWI inversion is performed using noisy multi-component seismic

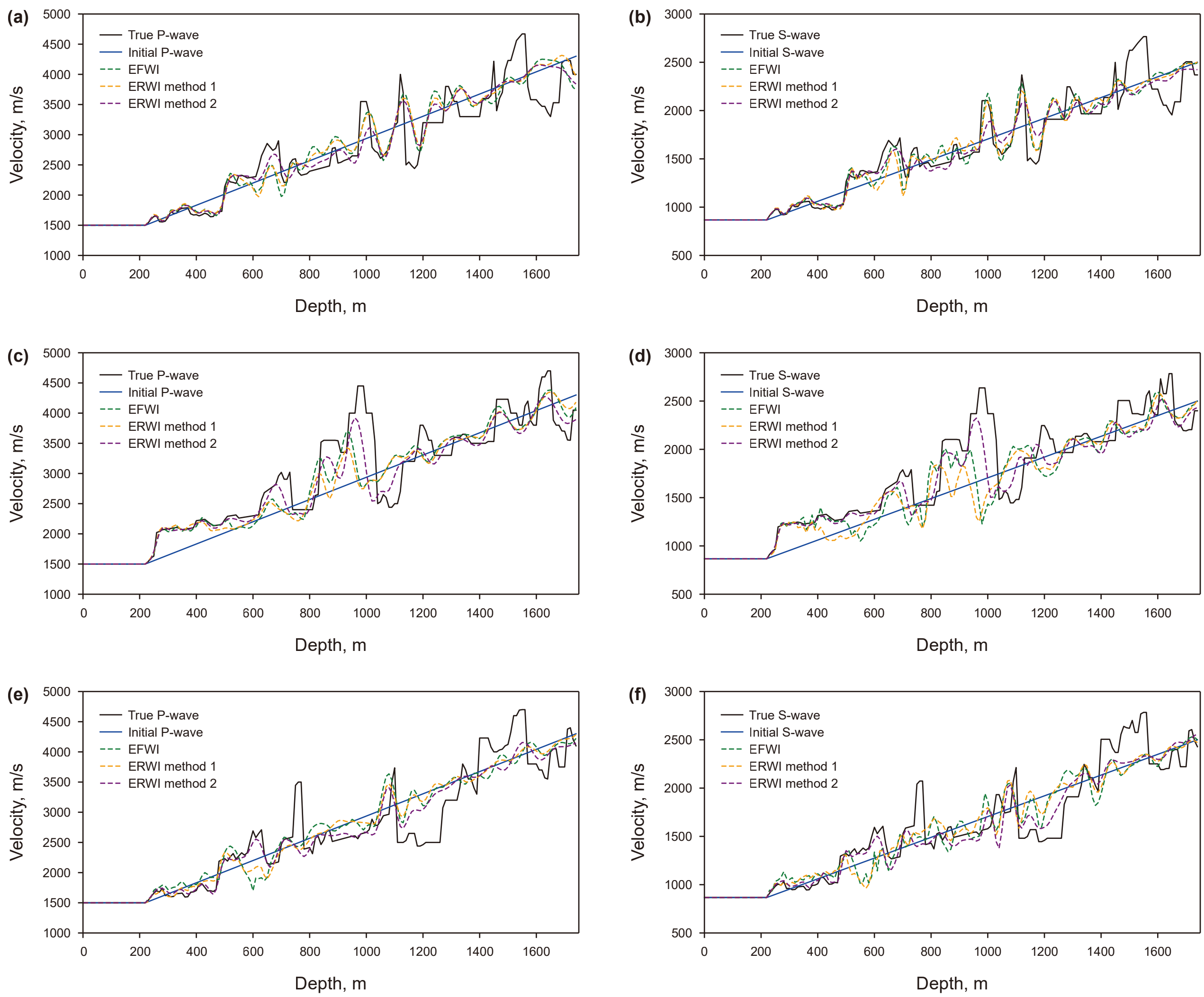

Fig. 17. The vertical profiles of **(a)** P-wave and **(b)** S-wave velocities at a horizontal distance of 0.6 km, **(c)** P-wave and **(d)** S-wave velocities at a horizontal distance of 1.2 km and **(e)** P-wave and **(f)** S-wave velocities at a horizontal distance of 1.8 km.

Table 1
The model errors of the inversion results of Marmousi2 model.

Parameter	Initial	EFWI	ERWI method 1	ERWI method 2
v_p	12.20%	12.33%	12.01%	10.42%
v_s	11.87%	12.33%	12.00%	10.16%

records, generated by adding Gaussian noise to the seismic data from the previous section. Considering all sources, receivers, time samples, and both horizontal and vertical components, the root-mean-square (RMS) of the seismic data is 4.61×10^{-10}, defined as

$$\mathrm{RMS} = \sqrt{\frac{1}{2\ N_s N_r N_t} \sum_{c\in\{x,z\}} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{t=1}^{N_t} \left[d_c^{\mathrm{signal}}(s,r,t) \right]^2}, \tag{25}$$

where N_s, N_r, and N_t denote the numbers of shots, receivers, and time samples, respectively. The standard deviation of the noise is 2.89×10^{-10}, computed as

$$\sigma_{\mathrm{noise}} = \sqrt{\frac{1}{2\ N_s N_r N_t} \sum_{c\in\{x,z\}} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{t=1}^{N_t} \left[d_c^{\mathrm{noise}}(s,r,t) - \overline{\theta} \right]^2}, \tag{26}$$

where $\overline{\theta}$ denotes the mean value of the noise, 3.25×10^{-14}, which is much smaller than the RMS values of both the seismic data and the noise, indicating that the noise can be regarded as approximately zero-mean. The signal-to-noise ratio (SNR) is 4.02 dB, representing a relatively strong noise level that provides a challenging test for evaluating the robustness of the proposed method. The SNR is defined based on the energy ratio between the signal and the noise as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{E_{\mathrm{signal}}}{E_{\mathrm{noise}}}\right), \tag{27}$$

where the signal or noise energy is defined as

$$E = \sum_{c\in\{x,z\}} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{t=1}^{N_t} [d_c(s,r,t)]^2. \tag{28}$$

Fig. 18 presents a comparison between the noise-free and noisy multi-component seismic data for the 12th common-shot gather.

The convergence curve in Fig. 19 shows that, under noisy conditions, the proposed ERWI method fails to reduce the data misfit to a sufficiently low level. This is because the noise in the observed

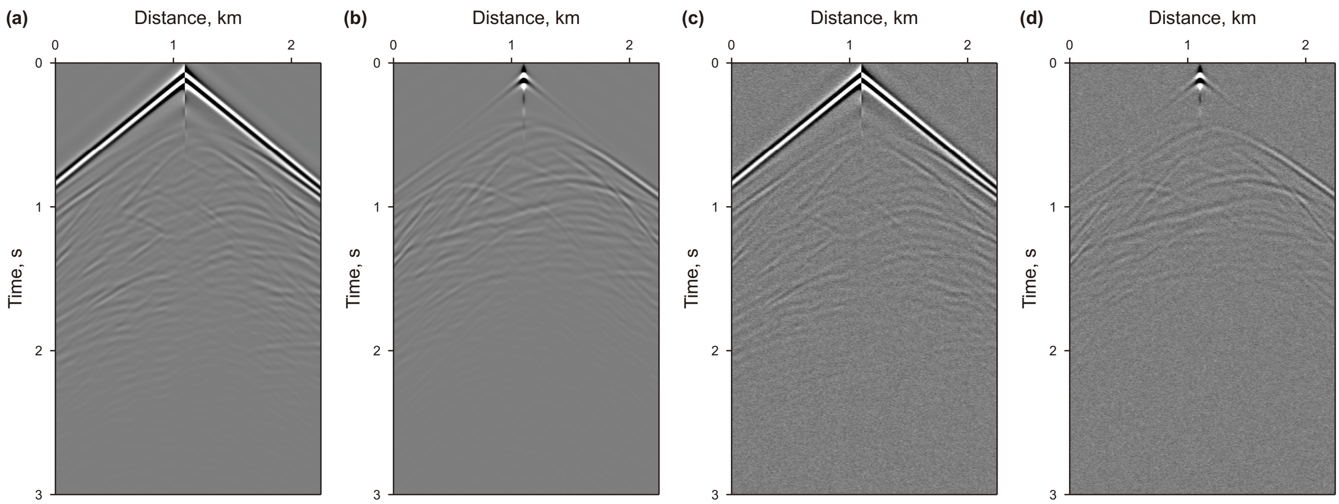

Fig. 18. The multi-component seismic records of the 12th shot for both noise-free and noisy data. **(a)** and **(b)** show the horizontal and vertical components of the noise-free seismic records, respectively, while **(c)** and **(d)** show the horizontal and vertical components of the noisy seismic records, respectively.

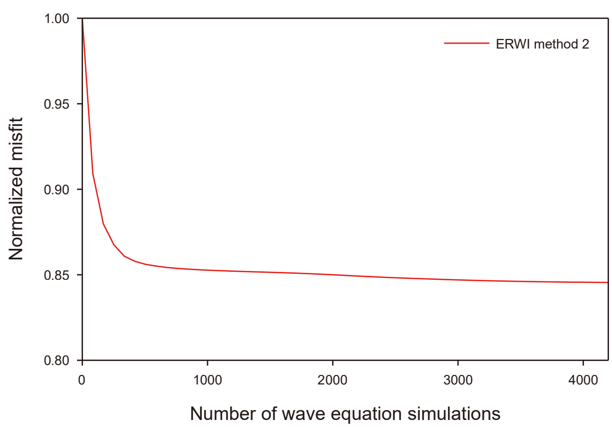

Fig. 19. The convergence curves of ERWI in the Marmousi2 model.

data cannot be effectively fitted. However, in inversion problems, the primary concern lies in the accuracy of the results in the model space. The final inversion results obtained from noisy data exhibit no significant degradation compared to the noise-free results shown in Fig. 14. Although the perturbation models in Fig. 14(c) and (d) are of higher quality than those in Fig. 20(c) and (d), the overall inversion results with noisy data remain acceptable, indicating that the proposed ERWI method exhibits robust performance under noisy conditions. In the case of noisy data, the model errors of the inverted P-wave and S-wave velocity models are 10.46% and 10.24%, respectively. Compared with the results obtained from noise-free data, these errors increase only slightly, demonstrating that the proposed method maintains strong robustness against noise.

4. Discussion

The successful implementation of FWI often relies on long-offset data, low-frequency information, and a good initial model. In the absence of these conditions, the inversion is prone to failure. Compared to FWI, RWI relaxes these requirements. By using deep penetrating reflected waves to update the background model, it can improve the inversion results. In RWI, the background and perturbation models can be explicitly separated, allowing low-wavenumber tomography components to be retrieved from the reflected waves. In the context of elastic media with multi-component data, ERWI simultaneously recovers both P-wave and S-wave velocities, thereby enriching the information of the subsurface parameters. Unlike a single reflected-wave objective function, the joint incident-wave and reflected-wave objective function leverages both transmitted and reflected waves to update the background model. The Fréchet derivatives and gradients of the background and perturbation models under this objective function share identical terms originating from transmitted waves propagating through the background model, whereas the perturbation model consists solely of these terms. Consequently, the perturbation model can be rationally integrated into the background model. Small-scale Gaussian smoothing is applied to ensure that the energy of the background wavefield remains concentrated in the transmitted waves.

During inversion, accounting for the Hessian matrix helps mitigate the effects of the acquisition system, geometrical spreading, and band-limited wavelets, while also reducing parameter trade-off effects. The accurate construction of the perturbation model is critical, as it not only influences the background model update but also directly integrates into the background model under the scale fusion strategy. To this end, a scheme is proposed to construct the perturbation model using the TGN method, implemented via an efficient Hessian-vector product and a matrix-free conjugate gradient approach. This method enhances the quality of the perturbation model and effectively suppresses multi-parameter trade-off effects. Under the proposed strategy, constructing the perturbation model using only the TGN method already yields significant improvements in inversion resolution, and the additional computational cost is manageable, especially considering that comparable computational effort produces better results. While the TGN method can also be applied to background model updates, doing so would further increase the computational burden compared to its use in perturbation model updates.

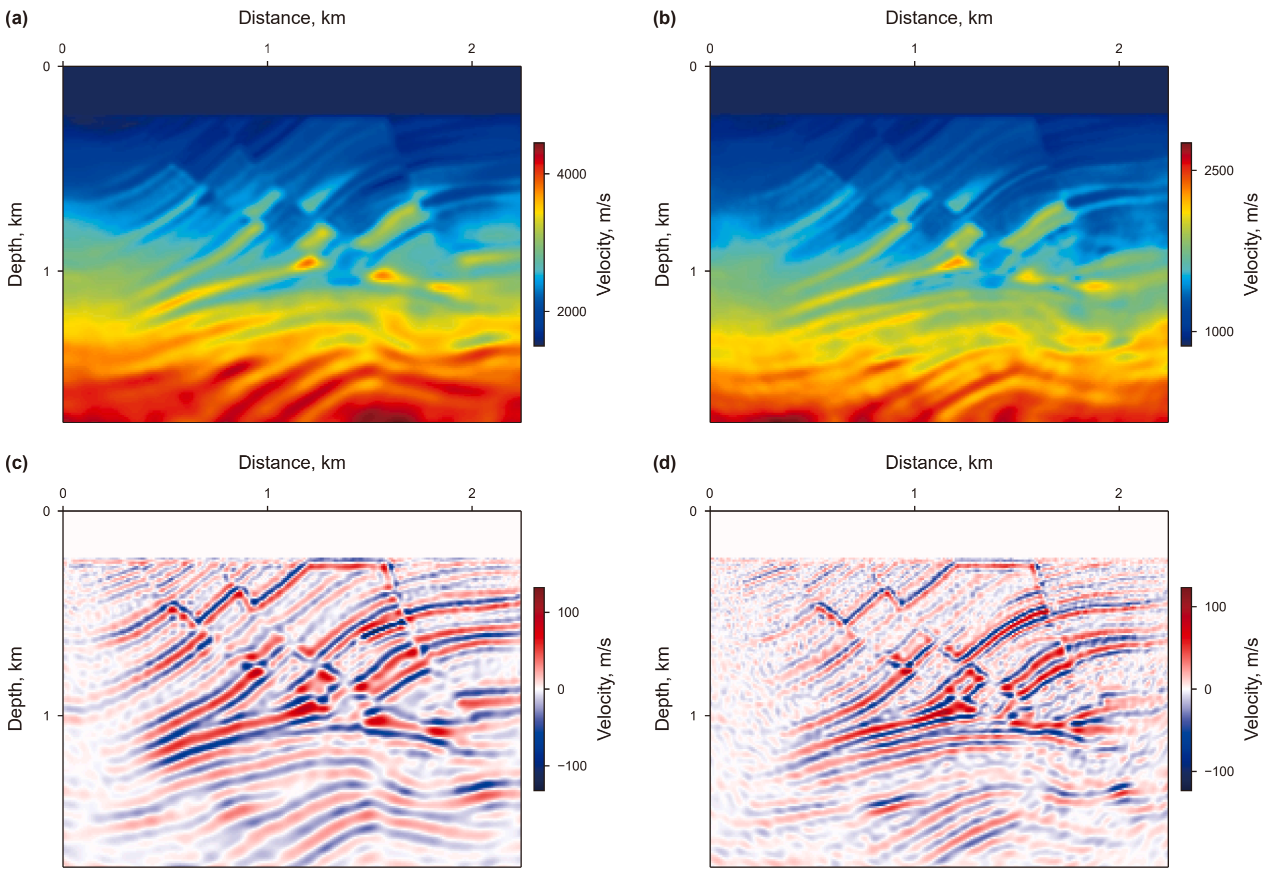

Fig. 20. The proposed ERWI inversion results of the Marmousi2 model using the proposed method. **(a)** and **(b)** show the P-wave and S-wave velocities of the background model, which represent the final inversion results. **(c)** and **(d)** show the P-wave and S-wave velocities of the temporary perturbation model at the final iteration.

5. Conclusion

We propose an ERWI method based on the Hessian and a scale fusion strategy. Within the joint incident and reflected wave objective function, the Fréchet derivatives and gradients of different wavefield components are derived and analyzed, clarifying the respective contributions of transmitted and reflected waves to the model updates. The update of the background model relies on the reflected waves generated by the perturbation model, emphasizing the importance of constructing accurate perturbation models and motivating the use of the Hessian matrix to enhance their inversion quality. Moreover, the shared terms in the Fréchet derivatives of the background and perturbation models inspire the adoption of a scale fusion strategy to improve the inversion process.

Compared with conventional ERWI, the proposed method employs the TGN method to construct more accurate perturbation models, which facilitates more reliable background model updates. The scale fusion strategy ensures that perturbation models are effectively integrated while their accuracy improves during iterations, establishing a mutually reinforcing feedback mechanism between the perturbation and background models. Numerical experiments on an anomaly model demonstrate the role of the Hessian in mitigating multi-parameter crosstalk and enhancing inversion accuracy. Tests using the Marmousi2 model indicate that, compared with EFWI and conventional ERWI, the proposed ERWI method achieves faster convergence, lower data misfit, and reduced model errors, resulting in significantly improved inversion results. Finally, noise-resilience tests confirm the robustness of the proposed method, highlighting its potential for future application to field data.

CRediT authorship contribution statement

Ming-Qian Wang: Writing – original draft, Software, Methodology, Conceptualization. **Bing-Shou He:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used ChatGPT in order to polish the English language and eliminate grammatical errors. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 42274149, and in part by the Natural Science Foundation of Shandong Province of China under Grant ZR2022MD074.

Appendix A. Fréchet derivatives

The forward modeling of the background wavefield is expressed as

$$\mathbf{F}(\mathbf{m}_0)\mathbf{u}_0 = \mathbf{f}. \tag{A.1}$$

Taking the partial derivative of Eq. (A.1) with respect to λ_0 and μ_0 yields

$$\frac{\partial \mathbf{F}(\mathbf{m}_0)}{\partial \lambda_0}\mathbf{u}_0 + \mathbf{F}(\mathbf{m}_0)\frac{\partial \mathbf{u}_0}{\partial \lambda_0} = 0, \tag{A.2a}$$

$$\frac{\partial \mathbf{F}(\mathbf{m}_0)}{\partial \mu_0}\mathbf{u}_0 + \mathbf{F}(\mathbf{m}_0)\frac{\partial \mathbf{u}_0}{\partial \mu_0} = 0. \tag{A.2b}$$

By incorporating the differential operator defined in Eq. (7)

$$\frac{\partial \mathbf{F}(\mathbf{m}_0)}{\partial \lambda_0} = -\mathbf{A}_1\mathbf{A}_\lambda, \tag{A.3a}$$

$$\frac{\partial \mathbf{F}(\mathbf{m}_0)}{\partial \mu_0} = -(\mathbf{A}_1\mathbf{A}_\mu + \mathbf{A}_2\mathbf{A}_\mu'), \tag{A.3b}$$

and utilizing the Green's function, Eq. (A.2) can be further expressed as

$$\frac{\partial \mathbf{u}_0}{\partial \lambda_0} = \mathbf{G}_0\mathbf{A}_1\mathbf{A}_\lambda\mathbf{u}_0, \tag{A.4a}$$

$$\frac{\partial \mathbf{u}_0}{\partial \mu_0} = \mathbf{G}_0\left(\mathbf{A}_1\mathbf{A}_\mu + \mathbf{A}_2\mathbf{A}_\mu'\right)\mathbf{u}_0. \tag{A.4b}$$

where $\mathbf{G}_0$ denotes the background Green's function.

By taking partial derivatives of the perturbed wavefield forward modeling Eq. (5) with respect to λ_0 and μ_0, respectively, we obtain

$$\frac{\partial \mathbf{F}(\mathbf{m}_0)}{\partial \lambda_0}\delta\mathbf{u} + \mathbf{F}(\mathbf{m}_0)\frac{\partial \delta\mathbf{u}}{\partial \lambda_0} = \mathbf{A}_1\delta\lambda\mathbf{A}_\lambda\frac{\partial \mathbf{u}_0}{\partial \lambda_0}, \tag{A.5a}$$

$$\frac{\partial \mathbf{F}(\mathbf{m}_0)}{\partial \mu_0}\delta\mathbf{u} + \mathbf{F}(\mathbf{m}_0)\frac{\partial \delta\mathbf{u}}{\partial \mu_0} = (\mathbf{A}_1\delta\mu\mathbf{A}_\mu + \mathbf{A}_2\delta\mu\mathbf{A}_\mu')\frac{\partial \mathbf{u}_0}{\partial \mu_0} \tag{A.5b}$$

Substituting Eq. (A.4) into Eq. (A.5) yields

$$\frac{\partial \delta\mathbf{u}}{\partial \lambda_0} = \mathbf{G}_0\mathbf{A}_1\delta\lambda\mathbf{A}_\lambda\mathbf{G}_0\mathbf{A}_1\mathbf{A}_\lambda\mathbf{u}_0 + \mathbf{G}_0\mathbf{A}_1\mathbf{A}_\lambda\delta\mathbf{u}, \tag{A.6a}$$

$$\begin{aligned}\frac{\partial \delta\mathbf{u}}{\partial \mu_0} &= \mathbf{G}_0\left(\mathbf{A}_1\delta\mu\mathbf{A}_\mu + \mathbf{A}_2\delta\mu\mathbf{A}_\mu'\right)\mathbf{G}_0\left(\mathbf{A}_1\mathbf{A}_\mu + \mathbf{A}_2\mathbf{A}_\mu'\right)\mathbf{u}_0 \\ &\quad + \mathbf{G}_0\left(\mathbf{A}_1\mathbf{A}_\mu + \mathbf{A}_2\mathbf{A}_\mu'\right)\delta\mathbf{u}\end{aligned} \tag{A.6b}$$

The perturbed Green's function satisfies Eq. (5) and can be expressed as

$$\mathbf{F}(\mathbf{m}_0)\delta\mathbf{G} = \mathbf{A}_1\delta\lambda\mathbf{A}_\lambda\mathbf{G}_0 + \mathbf{A}_1\delta\mu\mathbf{A}_\mu\mathbf{G}_0 + \mathbf{A}_2\delta\mu\mathbf{A}_\mu'\mathbf{G}_0 \tag{A.7}$$

Thus, the perturbed Green's function can be decomposed as $\delta\mathbf{G} = \delta\mathbf{G}_\lambda + \delta\mathbf{G}_\mu$.

$$\delta\mathbf{G}_\lambda = \mathbf{G}_0\mathbf{A}_1\delta\lambda\mathbf{A}_\lambda\mathbf{G}_0, \tag{A.8a}$$

$$\delta\mathbf{G}_\mu = \mathbf{G}_0(\mathbf{A}_1\delta\mu\mathbf{A}_\mu + \mathbf{A}_2\delta\mu\mathbf{A}_\mu')\mathbf{G}_0. \tag{A.8b}$$

Substituting Eq. (A.8) into Eq. (A.6) yields

$$\frac{\partial \delta\mathbf{u}}{\partial \lambda_0} = \delta\mathbf{G}_\lambda\mathbf{A}_1\mathbf{A}_\lambda\mathbf{u}_0 + \mathbf{G}_0\mathbf{A}_1\mathbf{A}_\lambda\delta\mathbf{u}, \tag{A.9a}$$

$$\frac{\partial \delta\mathbf{u}}{\partial \mu_0} = \delta\mathbf{G}_\mu(\mathbf{A}_1\mathbf{A}_\mu + \mathbf{A}_2\mathbf{A}_\mu')\mathbf{u}_0 + \mathbf{G}_0(\mathbf{A}_1\mathbf{A}_\mu + \mathbf{A}_2\mathbf{A}_\mu')\delta\mathbf{u} \tag{A.9b}$$

By combining Eqs. (A.4) and (A.9), the Fréchet derivatives with respect to the background model parameters are obtained as

$$\mathbf{L}_{\lambda_0} = \frac{\partial \mathbf{u}_0}{\partial \lambda_0} + \frac{\partial \delta\mathbf{u}}{\partial \lambda_0}, \tag{A.10a}$$

$$\mathbf{L}_{\mu_0} = \frac{\partial \mathbf{u}_0}{\partial \mu_0} + \frac{\partial \delta\mathbf{u}}{\partial \mu_0}. \tag{A.10b}$$

Taking the partial derivatives of the perturbed wavefield forward Eq. (5) with respect to $\delta\lambda$ and $\delta\mu$ and utilizing the Green's function, we obtain:

$$\frac{\partial \delta\mathbf{u}}{\partial \delta\lambda} = \mathbf{G}_0\mathbf{A}_1\mathbf{A}_\lambda\mathbf{u}_0, \tag{A.11a}$$

$$\frac{\partial \delta\mathbf{u}}{\partial \delta\mu} = \mathbf{G}_0(\mathbf{A}_1\delta\mu\mathbf{A}_\mu + \mathbf{A}_2\delta\mu\mathbf{A}_\mu')\mathbf{u}_0. \tag{A.11b}$$

It is generally assumed that the model perturbations do not influence the background wavefield. Therefore, the Fréchet derivatives with respect to the perturbation model parameters are given by

$$\mathbf{L}_{\delta\lambda} = \frac{\partial \delta\mathbf{u}}{\partial \delta\lambda}, \tag{A.12a}$$

$$\mathbf{L}_{\delta\mu} = \frac{\partial \delta\mathbf{u}}{\partial \delta\mu}. \tag{A.12b}$$

Appendix B. Gradient derivation using the adjoint-state method

By applying the Lagrange multiplier method and taking the forward equations of the background and perturbed wavefields as constraints, the optimization problem in Eq. (8) can be reformulated as

$$\begin{aligned}E(\mathbf{m}_0, \delta\mathbf{m}, \mathbf{u}_0, \delta\mathbf{u}, \boldsymbol{\Phi}, \boldsymbol{\Phi}') &= \frac{1}{2}\|\mathbf{u}_0 + \delta\mathbf{u} - \mathbf{u}_{\mathrm{obs}}\|_2^2 \\ &\quad + \langle \mathbf{F}(\mathbf{m}_0)\mathbf{u}_0 - \mathbf{f}, \boldsymbol{\Phi}' \rangle + \langle \mathbf{F}(\mathbf{m}_0)\delta\mathbf{u} - \mathbf{B}(\delta\mathbf{m})\mathbf{u}_0, \boldsymbol{\Phi} \rangle.\end{aligned} \tag{B.1a}$$

Here, $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}'$ denote the adjoint-state variables. The adjoint equations can be obtained by taking the derivative of Eq. (B.1) with respect to the state variables and setting it equal to zero:

$$\frac{\partial E}{\partial \mathbf{u}_0} = 0, \quad \frac{\partial E}{\partial \delta\mathbf{u}} = 0. \tag{B.2}$$

Thus, the adjoint equations can be written as

$$\mathbf{F}^{\mathrm{T}}(\mathbf{m}_0)\boldsymbol{\Phi} = -\Delta\mathbf{u}, \tag{B.3a}$$

$$\mathbf{F}^{\mathrm{T}}(\mathbf{m}_0)\boldsymbol{\Phi}' = -\Delta\mathbf{u} + \mathbf{B}^{\mathrm{T}}(\delta\mathbf{m})\boldsymbol{\Phi}. \tag{B.3b}$$

where T denotes the conjugate transpose operator, and $\Delta\mathbf{u} = \mathbf{u}_0 + \delta\mathbf{u} - \mathbf{u}_{\mathrm{obs}}$ represents the data residual. Introducing an auxiliary adjoint-state variable $\delta\mathbf{\Phi}$ and let

$$\mathbf{\Phi}' = \mathbf{\Phi} + \delta\mathbf{\Phi}. \tag{B.4}$$

Then subtracting Eq. (B.3a) from Eq. (B.3b), the adjoint equation can be further written as

$$\mathbf{F}^{\mathrm{T}}(\mathbf{m}_0)\mathbf{\Phi} = -\Delta\mathbf{u}, \tag{B.5a}$$

$$\mathbf{F}^{\mathrm{T}}(\mathbf{m}_0)\delta\mathbf{\Phi} = \mathbf{B}^{\mathrm{T}}(\delta\mathbf{m})\mathbf{\Phi}. \tag{B.5b}$$

Taking the derivative of Eq. (B.1) with respect to perturbation model parameters yields the gradients of the perturbation models:

$$\mathbf{g}(\delta\mathbf{m}) = \frac{\partial E}{\partial \delta\mathbf{m}} = -\left(\frac{\mathbf{B}(\delta\mathbf{m})}{\delta\mathbf{m}}\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{\Phi}. \tag{B.6}$$

The gradients with respect to the perturbation Lamé parameters can be written as

$$\frac{\partial E}{\partial \delta\lambda} = -(\mathbf{A}_1\mathbf{A}_\lambda\mathbf{u}_0)^{\mathrm{T}}\mathbf{\Phi}, \tag{B.7a}$$

$$\frac{\partial E}{\partial \delta\mu} = -\left(\mathbf{A}_1\mathbf{A}_\mu\mathbf{u}_0 + \mathbf{A}_2\mathbf{A}'_\mu\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{\Phi}, \tag{B.7b}$$

Expanding Eq. (B.7) further yields

$$\frac{\partial E}{\partial \delta\lambda} = -(\mathbf{A}_\lambda\mathbf{u}_0)^{\mathrm{T}}\mathbf{A}_1^{\mathrm{T}}\mathbf{\Phi}, \tag{B.8a}$$

$$\frac{\partial E}{\partial \delta\mu} = -(\mathbf{A}_\mu\mathbf{u}_0)^{\mathrm{T}}\mathbf{A}_1^{\mathrm{T}}\mathbf{\Phi} - \left(\mathbf{A}'_\mu\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{A}_2^{\mathrm{T}}\mathbf{\Phi}, \tag{B.8b}$$

Taking the derivative of Eq. (B.1) withrespect to background model parameters:

$$\frac{\partial E}{\partial \lambda_0} = \left(\frac{\partial\mathbf{F}(\mathbf{m}_0)}{\partial\lambda_0}\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{\Phi}' + \left(\frac{\partial\mathbf{F}(\mathbf{m}_0)}{\partial\lambda_0}\delta\mathbf{u}\right)^{\mathrm{T}}\mathbf{\Phi}, \tag{B.9a}$$

$$\frac{\partial E}{\partial \mu_0} = \left(\frac{\partial\mathbf{F}(\mathbf{m}_0)}{\partial\mu_0}\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{\Phi}' + \left(\frac{\partial\mathbf{F}(\mathbf{m}_0)}{\partial\mu_0}\delta\mathbf{u}\right)^{\mathrm{T}}\mathbf{\Phi}, \tag{B.9b}$$

Substituting Eqs. (A.3) and (B.4) into Eq. (B.9) yields

$$\frac{\partial E}{\partial \lambda_0} = -(\mathbf{A}_\lambda\mathbf{u}_0)^{\mathrm{T}}\mathbf{A}_1^{\mathrm{T}}\mathbf{\Phi} - (\mathbf{A}_\lambda\mathbf{u}_0)^{\mathrm{T}}\mathbf{A}_1^{\mathrm{T}}\delta\mathbf{\Phi} - (\mathbf{A}_\lambda\delta\mathbf{u})^{\mathrm{T}}\mathbf{A}_1^{\mathrm{T}}\mathbf{\Phi}, \tag{B.10a}$$

$$\begin{aligned}\frac{\partial E}{\partial \mu_0} &= -(\mathbf{A}_\mu\mathbf{u}_0)^{\mathrm{T}}\mathbf{A}_1^{\mathrm{T}}\mathbf{\Phi} - \left(\mathbf{A}'_\mu\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{A}_2^{\mathrm{T}}\mathbf{\Phi} \\ &\quad -(\mathbf{A}_\mu\mathbf{u}_0)^{\mathrm{T}}\mathbf{A}_1^{\mathrm{T}}\delta\mathbf{\Phi} - \left(\mathbf{A}'_\mu\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{A}_2^{\mathrm{T}}\delta\mathbf{\Phi} \\ &\quad -(\mathbf{A}_\mu\delta\mathbf{u})^{\mathrm{T}}\mathbf{A}_1^{\mathrm{T}}\mathbf{\Phi} - \left(\mathbf{A}'_\mu\delta\mathbf{u}\right)^{\mathrm{T}}\mathbf{A}_2^{\mathrm{T}}\mathbf{\Phi}.\end{aligned} \tag{B.10b}$$

Since seismic data consist of multiple shots and have a finite bandwidth, the final gradients of both the background and perturbation models are computed by integrating over all frequencies (or time) and summing the contributions from all shots.

$$\mathbf{g}(\lambda_0) = \sum_s \int \frac{\partial E}{\partial\lambda_0}\mathrm{d}\omega, \quad \mathbf{g}(\mu_0) = \sum_s \int \frac{\partial E}{\partial\mu_0}\mathrm{d}\omega, \tag{B.11a}$$

$$\mathbf{g}(\delta\lambda) = \sum_s \int \frac{\partial E}{\partial\delta\lambda}\mathrm{d}\omega, \quad \mathbf{g}(\delta\mu) = \sum_s \int \frac{\partial E}{\partial\delta\mu}\mathrm{d}\omega. \tag{B.11b}$$

Appendix C. Hessian-vector product

To compute the Hessian-vector product with respect to the perturbation model parameters, the second-order adjoint-state method (Wang et al., 2021; Assis et al., 2024; Wu et al., 2024) is employed to construct the Lagrangian function:

$$\begin{aligned}&P(\mathbf{g}(\delta\mathbf{m}), \mathbf{m}_0, \delta\mathbf{m}, \mathbf{u}_0, \delta\mathbf{u}, \mathbf{\Phi}, \delta\mathbf{\Phi}, \mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3, \mathbf{n}_4, \mathbf{n}_5) = \langle\mathbf{g}(\delta\mathbf{m}), \mathbf{p}\rangle \\ &\quad + \left\langle \mathbf{g}(\delta\mathbf{m}) + \left(\frac{\partial\mathbf{B}(\delta\mathbf{m})}{\partial\delta\mathbf{m}}\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{\Phi}, \mathbf{n}_1 \right\rangle + \langle\mathbf{F}(\mathbf{m}_0)\mathbf{u}_0 - \mathbf{f}, \mathbf{n}_2\rangle \\ &\quad + \langle\mathbf{F}(\mathbf{m}_0)\delta\mathbf{u} - \mathbf{B}(\delta\mathbf{m})\mathbf{u}_0, \mathbf{n}_3\rangle + \langle\mathbf{F}^{\mathrm{T}}(\mathbf{m}_0)\mathbf{\Phi} + \Delta\mathbf{u}, \mathbf{n}_4\rangle \\ &\quad + \langle\mathbf{F}^{\mathrm{T}}(\mathbf{m}_0)\delta\mathbf{\Phi} - \mathbf{B}^{\mathrm{T}}(\delta\mathbf{m})\mathbf{\Phi}, \mathbf{n}_5\rangle.\end{aligned} \tag{C.1}$$

where $\mathbf{g}(\delta\mathbf{m})$ denotes the gradient with respect to the perturbation model parameters, $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$, $\mathbf{n}_4$, and $\mathbf{n}_5$ are the second-order adjoint-state variables.

Taking the derivative of Eq. (C.1) with respect to $\delta\mathbf{m}$ yields the Hessian-vector product:

$$\mathbf{H}_{\delta\mathbf{m}}\mathbf{p} = \frac{\partial\mathbf{g}(\delta\mathbf{m})}{\partial\delta\mathbf{m}}\mathbf{p} = -\left(\frac{\partial\mathbf{B}(\delta\mathbf{m})}{\partial\delta\mathbf{m}}\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{n}_3 - \left(\frac{\partial\mathbf{B}^{\mathrm{T}}(\delta\mathbf{m})}{\partial\delta\mathbf{m}}\mathbf{\Phi}\right)^{\mathrm{T}}\mathbf{n}_5. \tag{C.2}$$

Taking the derivative of Eq. (C.1) with respect to the state variables and setting it equal to zero:

$$\frac{\partial P}{\partial\mathbf{g}(\delta\mathbf{m})} = 0, \quad \frac{\partial P}{\partial\mathbf{u}_0} = 0, \quad \frac{\partial P}{\partial\delta\mathbf{u}} = 0, \quad \frac{\partial P}{\partial\mathbf{\Phi}} = 0, \quad \frac{\partial P}{\partial\delta\mathbf{\Phi}} = 0. \tag{C.3}$$

The second-order adjoint equations can be obtained as

$$\mathbf{p} + \mathbf{n}_1 = 0, \tag{C.4a}$$

$$\mathbf{F}(\mathbf{m}_0)\mathbf{n}_5 = 0, \tag{C.4b}$$

$$\mathbf{F}(\mathbf{m}_0)\mathbf{n}_4 = \mathbf{B}(\delta\mathbf{m})\mathbf{n}_5 - \frac{\partial\mathbf{B}(\delta\mathbf{m})}{\partial\delta\mathbf{m}}\mathbf{u}_0\mathbf{n}_1, \tag{C.4c}$$

$$\mathbf{F}^{\mathrm{T}}(\mathbf{m}_0)\mathbf{n}_3 = -\mathbf{n}_4, \tag{C.4d}$$

$$\mathbf{F}^{\mathrm{T}}(\mathbf{m}_0)\mathbf{n}_2 = \mathbf{B}^{\mathrm{T}}(\delta\mathbf{m})\mathbf{n}_3 - \left(\left(\frac{\partial\mathbf{B}(\delta\mathbf{m})}{\delta\mathbf{m}}\right)^{\mathrm{T}}\mathbf{\Phi}\right)^{\mathrm{T}}\mathbf{n}_1, \tag{C.4e}$$

By neglecting the terms irrelevant to Eq. (C.2) and further simplifying, we obtain

$$\mathbf{p} + \mathbf{n}_1 = 0, \tag{C.5a}$$

$$\mathbf{F}(\mathbf{m}_0)\mathbf{n}_4 = -\mathbf{B}(\mathbf{n}_1)\mathbf{u}_0, \tag{C.5b}$$

$$\mathbf{F}^{\mathrm{T}}(\mathbf{m}_0)\mathbf{n}_3 = -\mathbf{n}_4, \tag{C.5c}$$

and

$$\mathbf{H}^{a}_{\delta\mathbf{m}}\mathbf{p} = -\left(\frac{\partial\mathbf{B}(\delta\mathbf{m})}{\partial\delta\mathbf{m}}\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{n}_3. \tag{C.6}$$

References

Aki, K., Richards, P.G., 2002. Quantitative Seismology. University Science Books, Sausalito, CA.
Alkhalifah, T., 2014. Scattering-angle based filtering of the waveform inversion gradients. Geophys. J. Int. 200, 363–373. https://doi.org/10.1093/gji/ggu379.
Assis, C.A.M., Chauris, H., Audebert, F., Williamson, P., 2024. Investigating hessian-based inversion velocity analysis. Geophysics 89, R169–R186. https://doi.org/10.1190/geo2022-0689.1.
Brenders, A.J., Pratt, R.G., 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., 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.
Bunks, C., Saleck, F.M., Zaleski, S., Chavent, G., 1995. Multiscale seismic waveform inversion. Geophysics 60, 1457–1473. https://doi.org/10.1190/1.1443880.
Chen, K., Sacchi, M.D., 2020. Time-domain elastic Gauss–Newton full-waveform inversion: A matrix-free approach. Geophys. J. Int. 223, 1007–1039. https://doi.org/10.1093/gji/ggaa330.
Chi, B.X., Dong, L.G., Liu, Y.Z., 2014. Full waveform inversion method using envelope objective function without low frequency data. J. Appl. Geophys. 109, 36–46. https://doi.org/10.1016/j.jappgeo.2014.07.010.
Da Silva, S., Karsou, A., de Souza, A., Capuzzo, F., Costa, F., Moreira, R., Cetale, M., 2022. A graph-space optimal transport objective function based on q-statistics to mitigate cycle-skipping issues in FWI. Geophys. J. Int. 231 (2), 1363–1385. https://doi.org/10.1093/gji/ggac267.
Eisenstat, S.C., Walker, H.F., 1996. Choosing the forcing terms in an inexact newton method. SIAM J. Sci. Comput. 17, 16–32. https://doi.org/10.1137/0917003.
Eslaminia, M., Elmeliegy, A.M., Guddati, M.N., 2022. Full waveform inversion through double-sweeping solver. J. Comput. Phys. 453, 110914. https://doi.org/10.1016/j.jcp.2021.110914.
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.
Guo, Q., Alkhalifah, T., 2017. Elastic reflection-based waveform inversion with a nonlinear approach. Geophysics 82, R309–R321. https://doi.org/10.1190/geo2016-0407.1.
Ha, W., Kang, S.G., Shin, C., 2015. 3d laplace-domain waveform inversion using a low-frequency time-domain modeling algorithm. Geophysics 80, R1–R13. https://doi.org/10.1190/geo2013-0332.1.
Huang, G.H., Nammour, R., Symes, W.W., 2018. Volume source-based extended waveform inversion. Geophysics 83, R369–R387. https://doi.org/10.1190/geo2017-0330.1.
Lailly, P., Bednar, J., 1983. The seismic inverse problem as a sequence of before stack migrations. In: Conference on Inverse Scattering: Theory and Application. Society for Industrial and Applied Mathematics, Philadelphia, PA, pp. 206–220.
Li, Y.Y., Guo, Q., Li, Z.C., Alkhalifah, T., 2019. Elastic reflection waveform inversion with variable density. Geophysics 84, R553–R567. https://doi.org/10.1190/geo2017-0722.1.
Liu, Y.Z., Yang, J.Z., Chi, B.X., Dong, L.G., 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.
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.
Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., Virieux, J., 2016. Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion. Geophys. J. Int. 205, 345–377. https://doi.org/10.1093/gji/ggw014.
Métivier, L., Brossier, R., Operto, S., Virieux, J., 2017. Full waveform inversion and the truncated newton method. SIAM Rev. 59, 153–195. https://doi.org/10.1137/120877854.
Operto, S., Virieux, J., Dessa, J.X., Pascal, G., 2006. Crustal seismic imaging from multifold ocean bottom seismometer data by frequency domain full waveform tomography. J. Geophys. Res. Solid Earth 111. https://doi.org/10.1029/2005JB003835.
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, R49–R64. https://doi.org/10.1190/geo2015-0595.1.
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.
Prieux, V., Brossier, R., Operto, S., Virieux, J., 2013. Multiparameter full waveform inversion of multicomponent obc data from valhall. Part 2: Imaging compressional and shear-wave velocities. Geophys. J. Int. 194, 1665–1681. https://doi.org/10.1093/gji/ggt178.
Ren, Z.M., Liu, Y., 2016. A hierarchical elastic full-waveform inversion scheme based on wavefield separation and the multistep-length approach. Geophysics 81, R99–R123. https://doi.org/10.1190/geo2015-0431.1.
Saad, Y., 2003. Iterative Methods for Sparse Linear Systems. Society for Industrial and Applied Mathematics, Philadelphia, PA.
Shin, C.S., Cha, Y.H., 2008. Waveform inversion in the laplace domain. Geophys. J. Int. 173, 922–931. https://doi.org/10.1111/j.1365-246X.2008.03768.x.
Sirgue, L., Barkved, O.I., Dellinger, J., Etgen, J., Albertin, U., Kommedal, J.H., 2010. Thematic set: Full waveform inversion: the next leap forward in imaging at Valhall. First Break 28. https://doi.org/10.3997/1365-2397.2010012.
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.
Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation. Geophysics 49, 1259–1266. https://doi.org/10.1190/1.1441754.
Van Leeuwen, T., Mulder, W.A., 2010. A correlation-based misfit criterion for wave-equation traveltime tomography. Geophys. J. Int. 182, 1383–1394. https://doi.org/10.1111/j.1365-246X.2010.04681.x.
Vigh, D., Jiao, K., Watts, D., Sun, D., 2014. Elastic full-waveform inversion application using multicomponent measurements of seismic data collection. Geophysics 79, R63–R77. https://doi.org/10.1190/geo2013-0055.1.
Vigh, D., Starr, E.W., 2008. 3D prestack plane-wave, full-waveform inversion. Geophysics 73, VE135–VE144. https://doi.org/10.1190/1.2952623.
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., Geng, J.H., 2021. Reflection full waveform inversion with second-order optimization using the adjoint-state method. J. Geophys. Res. Solid Earth 126. https://doi.org/10.1029/2021JB022135 e2021JB022135.
Wang, T.F., Cheng, J.B., Geng, J.H., 2022. Reflection-based traveltime and waveform inversion with second-order optimization. Pet. Sci. 19, 1582–1591. https://doi.org/10.1016/j.petsci.2022.02.003.
Warner, M., Guasch, L., 2016. Adaptive waveform inversion: Theory. Geophysics 81, R429–R445. https://doi.org/10.1190/geo2015-0387.1.
Wu, R.S., Luo, J.R., Wu, B.Y., 2014. Seismic envelope inversion and modulation signal model. Geophysics 79, WA13–WA24. https://doi.org/10.1190/geo2013-0294.1.
Wu, S.H., Wang, T.F., Cheng, J.B., 2024. Second-order optimization for multiparameter reflection waveform inversion in acoustic VTI media. Geophys. J. Int. 236, 249–269. https://doi.org/10.1093/gji/ggad406.
Wu, Z., 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.
Xie, C., Qin, Z.L., Wang, J.H., Song, P., Shen, H.G., Yu, S.Q., Ma, B.J., Liu, X.Q., 2024. Full waveform inversion based on hybrid gradient. Pet. Sci. 21 (3), 1660–1670. https://doi.org/10.1016/j.petsci.2024.01.013.
Xing, G.C., Zhu, T.Y., 2024. Advancing attenuation estimation through integration of the hessian in multiparameter viscoacoustic full-waveform inversion. Geophysics 89, 1–55. https://doi.org/10.1190/geo2023-0634.1.
Xu, S., Chen, F., Lambaré, G., Zhang, Y., Wang, D.L., 2012. Inversion on reflected seismic wave. In: SEG Technical Program Expanded Abstracts 2012. Society of Exploration Geophysicists, Las Vegas, NV, pp. 1–5.
Xu, W.C., Wang, T.F., Cheng, J.B., 2019. Elastic model low-to intermediate-wavenumber inversion using reflection traveltime and waveform of multicomponent seismic data. Geophysics 84, R123–R137. https://doi.org/10.1190/geo2018-0306.1.
Xu, W.C., Wang, T.F., Cheng, J.B., Geng, J.H., 2021. Hessian-based reflection waveform inversion: Theory and applications to synthetic and field data. Geophysics 86, R747–R762. https://doi.org/10.1190/geo2020-0601.1.
Yang, J.Z., Liu, Y.Z., Dong, L.G., 2016. Simultaneous estimation of velocity and density in acoustic multiparameter full-waveform inversion using an improved scattering-integral approach. Geophysics 81, R399–R415. https://doi.org/10.1190/geo2015-0707.1.
Yao, G., Da-Silva, N.V., Warner, M., Kalinicheva, T., 2018. Separation of migration and tomography modes of full-waveform inversion in the plane wave domain. J. Geophys. Res. Solid Earth 123, 1486–1501. https://doi.org/10.1002/2017JB015207.
Yao, G., Wu, D., Wang, S.X., 2020. A review on reflection-waveform inversion. Pet. Sci. 17, 334–351. https://doi.org/10.1007/s12182-020-00431-3.
Yong, P., Liao, W.Y., Huang, J.P., Li, Z.C., Lin, Y.T., 2019. Misfit function for full waveform inversion based on the wasserstein metric with dynamic formulation. J. Comput. Phys. 399, 108911. https://doi.org/10.1016/j.jcp.2019.108911.
Zhang, Q.C., Zhou, H., Li, Q.Q., Chen, H.M., Wang, J., 2016. Robust source-independent elastic full-waveform inversion in the time domain. Geophysics 81, R29–R44. https://doi.org/10.1190/geo2015-0073.1.
Zhou, W., Brossier, R., Operto, S., Virieux, J., 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.