Geophysical Journal International

Geophys. J. Int. (2024) **238,** 132–155
Advance Access publication 2024 April 18
GJI Seismology

https://doi.org/10.1093/gji/ggae143

Multiparameter shallow-seismic waveform inversion based on the Jensen–Shannon divergence

Yingwei Yan,[1,2,3] Xiaofei Chen,[2,3] Jing Li,[4] Jianbo Guan,[5] Yu Li[6] and Shihao Cui[7]

[1]*Department of Earth and Space Sciences, Southern University of Science and Technology, Shenzhen* 518055*, China*
[2]*Shenzhen Key Laboratory of Deep Offshore Oil and Gas Exploration Technology, Southern University of Science and Technology, Shenzhen* 518055*, China. E-mail: chenxf@sustech.edu.cn*
[3]*Guangdong Provincial Key Laboratory of Geophysical High-resolution Imaging Technology, Southern University of Science and Technology, Shenzhen* 518055*, China*
[4]*College of Geo-exploration Science and Technology, Jilin University, Changchun* 130015*, China*
[5]*School of Earth Sciences, Zhejiang University, Hangzhou* 310027*, China*
[6]*School of Geological Engineering and Geomatics, Chang'an University, Xi'an* 710054*, China*
[7]*Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, Montréal H3C3A7, Canada*

Accepted 2024 April 10. Received 2024 February 22; in original form 2023 July 13

SUMMARY

Seismic full-waveform inversion (FWI) or waveform inversion (WI) has gained extensive attention as a cutting-edge imaging method, which is expected to reveal the high-resolution images of complex geological structures. In this paper, we regard each 1-D signal in the inversion system as a 1-D probability distribution, then use the Jensen–Shannon divergence from information theory to measure the discrepancy between the predicted and observed signals, and finally implement a novel 2-D multiparameter shallow-seismic WI (MSWI). Essentially, the novel approach achieves an implicit weighting along the time-axis for each 1-D adjoint source defined by the classical WI (CWI), thus enhancing the extra illumination for a deeper medium compared with the CWI. By evaluating the inversion results of the two-layer model and fault model, the reconstruction accuracy for *S*-wave velocity and density of the new method is increased by about 30 and 20 per cent compared with that of the CWI under the same conditions, respectively. The reconstruction performance for *P*-wave velocity of these two methods is almost equal. In addition, the new 2-D MSWI is also resilient to white Gaussian noise in the data. Numerically, the inversion system has almost the strongest sensitivities to the *S*-wave velocity and density, performing the poorest sensitivity to the *P*-wave velocity. Finally, we test the novel method with a detection case for a power tunnel.

Key words: Inverse theory; Surface waves and free oscillations; Waveform inversion.

1 INTRODUCTION

An important aspect of geophysics is the inversion (Snieder & Trampert 1999), or imaging related to the parameter estimation (Mario & Patrizia 1998; Aster *et al.* 2018), where the seismic inversion plays a critical role in revealing the parameter distribution of underground medium (Wang 2017). The imaging methods based on the seismic waves may be divided into two categories: the body-wave (reflection seismology, first-arrival wave tomography, and so on) and surface-wave imaging (mainly utilizing the dispersion; Wang *et al.* 2023). Compared with the body waves, the surface waves (referred to Rayleigh waves) have the characteristics of large amplitude, dispersion and cylindrical-type geometrical spreading (Foti *et al.* 2014; Xia *et al.* 2015), making the surface-wave exploration is widely used in the geological structure survey with various scales (Xia *et al.* 2015; Chen *et al.* 2022; Yan *et al.* 2022a). The core of surface-wave exploration is the dispersion analysis, which often identifies the 1-D *S*-wave velocity (V_S) structures below the survey points through inverting the frequency-dependent phase-velocities (Xia *et al.* 2015; Foti *et al.* 2018; Vantassel & Cox 2021). Currently, the common inversion strategies often limit the attention to the fundamental-mode dispersion curve only because of the potential mode misidentification (Lamuraglia *et al.* 2023), although many researchers have emphasized that the higher-mode dispersion information is essential to better constrain the 1-D *S*-wave velocity solution (Feng *et al.* 2005; Luo *et al.* 2009; Farrugia *et al.* 2016; Yan *et al.* 2022b). This means that the multisolution degree of the common methods would be serious, for this, some new inversion strategies for the multimodal dispersion curves have been reported (Maraschini & Foti 2010; Yan *et al.* 2022b). However, the intrinsic 1-D limitation still remains.

The 1-D horizontal layered media assumption makes that the clas-

© The Author(s) 2024. Published by Oxford University Press on behalf of The Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

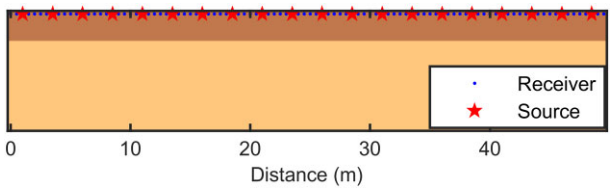

Figure 1. The schematic diagram of the observed system of the reconstruction tests for the two-layer model and fault model with noise-free data and poor initial models. Both the sources and receivers are configured on the free-surface, the red pentagrams indicate the location of the source on the free-surface for each shot gather, and the blue dots represent the positions of the corresponding receivers.

sic dispersion inversion is only suitable for the medium with low lateral heterogeneities (Gao *et al.* 2023), in other words, only an average V_S structure can be attained. To extract the lateral varying structure from the dispersion information, Hu *et al.* (2021) presented a multiscale window analysis of surface waves (MWASW), which establishes a forward algorithm of the fundamental-mode dispersion curve for 2-D model based on the plane-wave basic function, therefore the inversion is designed naturally. But this approach lacks the consideration for the higher-mode surface waves and scattered waves, possibly facing the dilemma for detecting the obstacles and sharp discontinuous interfaces. Most important of all, the dispersion curves are the sparse representations of waveform data, therefore, it would inevitably lose the information correlated to the medium.

The full-waveform inversion (FWI) fits the observed waveforms, directly (Tarantola 1984), without the complex dimensionality reduction and data transformation process, which have great potential to give the high-resolution parameter estimation for complex 2-D or 3-D geological structures (Gao *et al.* 2023). Under many circumstances, the real-amplitude information of the observed data related to the geometric diffusion and attenuation of the media is not utilized by the inversion system, and a trace-by-trace normalization strategy is adopted, which is named as the waveform inversion (WI) in the paper. Previous researchers paid great focus on the body-wave FWI or WI (Mora 1987; Crase *et al.* 1990; Vigh *et al.* 2014; Pan *et al.* 2016), which means that the application for field data needs to filter out the surface-wave components. However, the surface-wave components dominate in the near-surface seismic wavefield (Xia *et al.* 2015; Gao *et al.* 2023), therefore, the surface waves are indispensable for the near-surface FWI and WI. Due to the advancement of the seismometers and the accessibility of huge computing power, the FWI/WI has become a cutting-edge imaging technique for near-surface medium, and many field case studies have been reported (Romdhane *et al.* 2011; Tran *et al.* 2013; Groos *et al.* 2014, 2017; Li *et al.* 2017; Pan *et al.* 2019; Gao *et al.* 2020, 2023). It is worth mentioning that, compared with the FWI/WI driven by the 2-D wave equation of *P*-SV type (the observed data is dominated by the Rayleigh waves), the FWI/WI of the 2-D wave equation of SH type (the Love-wave is dominant of the observed data) has a simpler parameter classes, which has also received extensive attention (Dokter *et al.* 2017; Köhn *et al.* 2018; Guan *et al.* 2022), although it may be more difficult to excite the Love-wave with high signal-to-noise ratio by active sources.

As is known to all, the cycle skipping is a non-negligible issue for the body-wave FWI/WI (Engquist & Yang 2022), therefore, a favourable initial model is required for the classical approaches. However, from the stepwise refinement perspective, the cycle skipping is not very critical for the FWI/WI mainly driven by the surface waves because a nice initial model can be given by the dispersion spectrum analysis. Furthermore, it has been suggested that this initial model construction strategy is effective enough to avoid the cycle skipping (Guan *et al.* 2022; Gao *et al.* 2023). A more important aspect is the insufficient illumination for the deeper media because the receivers and sources are deployed on the free-surface for most FWI/WI cases, in contrast to the medical imaging, this is an observation system with very limited angle coverage. Therefore, there may be such a problem, how should we effectively measure the differences between the predicted and observed signals to mine the information as much as possible, so that achieving a high-precision imaging in the situation of insufficient observation?

Choi & Alkhalifah (2012) proposed to adopt the global correlation norm to measure the discrepancy between the observed and predicted waveforms to attain a better illumination, essentially, this misfit function realizes the reform for the adjoint wavefield generated by the L^2 norm. Wu *et al.* (2014) suggested the envelope-based misfit, and Borisov *et al.* (2018) extended the objective function to 3-D elastic surface-wave FWI for inverting the shallow large-scale heterogeneities. Liu *et al.* (2022) constructed a new misfit for seismic-wave WI through using the instantaneous phase of 1-D signal, which is considered to be able to uniformly extract the information from different events of seismic signals, and they tested the new misfit by an archaeogeophysics example. These researches are based on such a fact that the different misfit functions have different properties.

Inspired by this, we regard the signal of each seismic trace as a 1-D probability distribution, and measure the difference between the predicted and observed data sets from information theory (Shannon 1948). The difference between two probability distribution is often measured by the information entropy, and the Jensen–Shannon divergence (JSD) is an extension for the Shannon entropy (Lin 1991). Some studies have shown that the entropy is more suitable to estimate the complexity of the detail modes (the localized oscillation and scatter/wavelets) of the signal compared with the apparent fluctuations of the observed signal, and the detail modes of the signal usually carry a lot of information related to the complex system (Tavares & Lucena 2005). In other words, the JSD has the capacity to capture more information from the seismic signals, which is valuable for many seismic imaging problems. Therefore, we construct a novel 2-D multiparameter shallow-seismic WI (MSWI) by using the JSD (Manning & Schütze 1999) to measure the similarity between all the 1-D predicted and observed probability distributions. The JSD reforms the misfit function compared with that defined by the L^2 norm, and then generates the different adjoint wavefield and gradient vector, which is likely to produce the better descent directions for inversion due to its nice measurement ability of signal. The existed researches (Gao *et al.* 2020; Liu *et al.* 2021) pointed that the pure elastic setup is still applicable when the attenuation is weak or the Q model is about laterally homogeneous, and that, our adopted source correction filter also decreases the adverse effect because of the lack of Q model in the parametrization design. Therefore, we ignore the attenuation for our presented 1.0 version 2-D MSWI based on the JSD in the paper, only reconstruct the V_S, *P*-wave velocity (V_P) and density (ρ).

The rest of this paper is organized as follows. The methodology is introduced in Section 2, including the misfit function definition, calculation formula for the gradient vector and quasi-Hessian operator, and the preconditioned steepest-descent (PSD) for optimization that

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

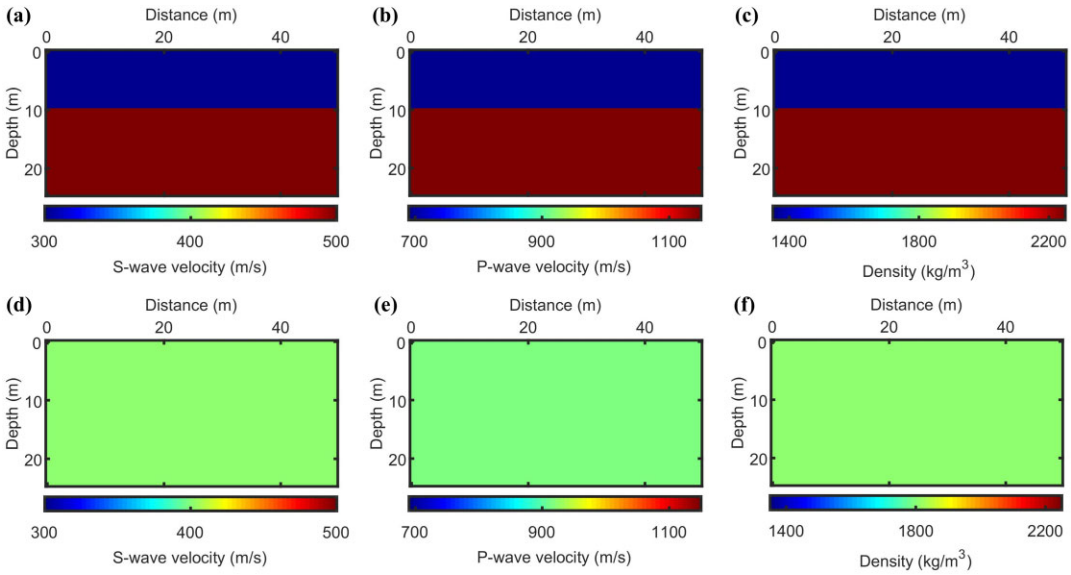

Figure 2. The true and initial models of the reconstruction test for the two-layer model. (a), (b) and (c) are the true V_S, V_P and ρ model, respectively. (d), (e) and (f) represent the initial V_S, V_P and ρ model, respectively.

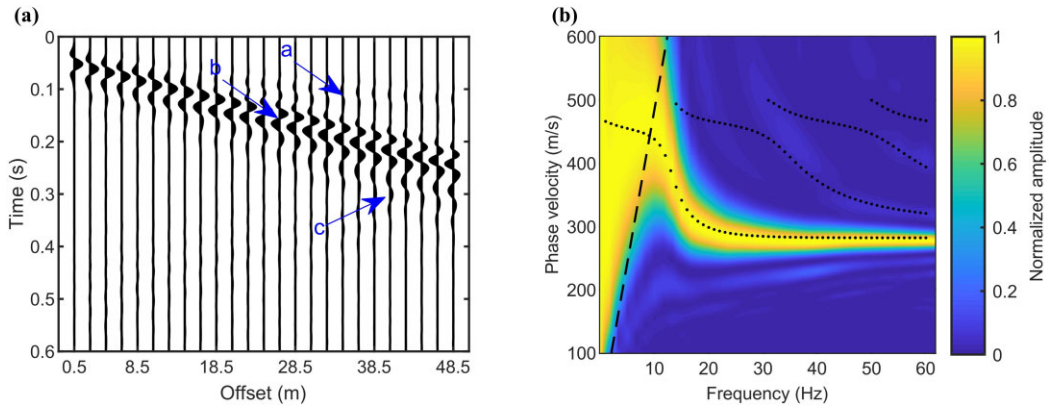

Figure 3. The normalized seismograms (a) for the first shot gather of the two-layer model, and the trace-spacing is adjusted as 2 m for clarity. The blue lowercases a, b and c with arrows in (a) indicate the refracted-wave, Rayleigh waves and scattered waves from the stratum interface, respectively. (b) is the dispersion spectrum of (a) calculated by the phase-shift method, the black dashed line is defined by the length of the receiver-array of the first shot gather, which constrains the maximum measurable wavelength of Rayleigh waves, and the black dotted lines denote the analytic phase velocities of the two-layer model calculated by the multimode fast vector transfer algorithm (Fan *et al.* 2007).

has nice convergence and iterative sustainability (Yan *et al.* 2022b). In Section 3, we first evaluate the performances of our presented method and the classical WI (CWI) through the reconstruction tests for two-layer model and fault model under the condition of poor initial models. Then, we consider the synthetic tests that are closer to the real-world situation, that is, the better initial models and the waveform data contains the white Gaussian noise. As well as, we also design a reverse test for the field data case by sharing the same observation system as the field data and deriving the test model according to the inverted results of the field data. The sensitivities of the inversion system to different parameters are also analysed in this part. In Section 4, we verify the proposed method through the seismic-wave data set collected above a power tunnel, and the lateral and vertical positions of the tunnel are successfully identified from the inverted results. Finally, we revisit the high-efficiency computation method for the gradient vector and quasi-Hessian operator by the Lagrange multiplier method (Bertsekas 1982; Plessix 2006) in Appendix A, and the source correction filter from a damped least-square perspective in Appendix B.

2 METHOD

2.1 The definition of misfit function

For the classical FWI, the misfit function is often derived by the L^2 norm. When there is only one shot gather in the inversion system, the misfit function is expressed as:

$$E_{L^2}(p) = \frac{1}{2}\|Lu(p) - d\|_2^2 = \frac{1}{2}\langle Lu(p) - d, Lu(p) - d\rangle . \quad (1)$$

Here, the symbol $||\ldots||_2$ represents the L^2 norm, and the above misfit function is defined as the square of the L^2 norm of the data residuals. Where L is the sampling operator to the receiver locations, $u(p)$ represents the seismic wavefield related to the physical parameter p, which is obtained by solving the 2-D wave equation of P-SV type of elastic media in the paper, d denotes the observed data, the parameter p is parametrized as the V_S, V_P and ρ in this paper, and the symbol $<\ldots,\ldots>$ denotes the inner product operation. For CWI, a trace-by-trace normalization strategy is adopted

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

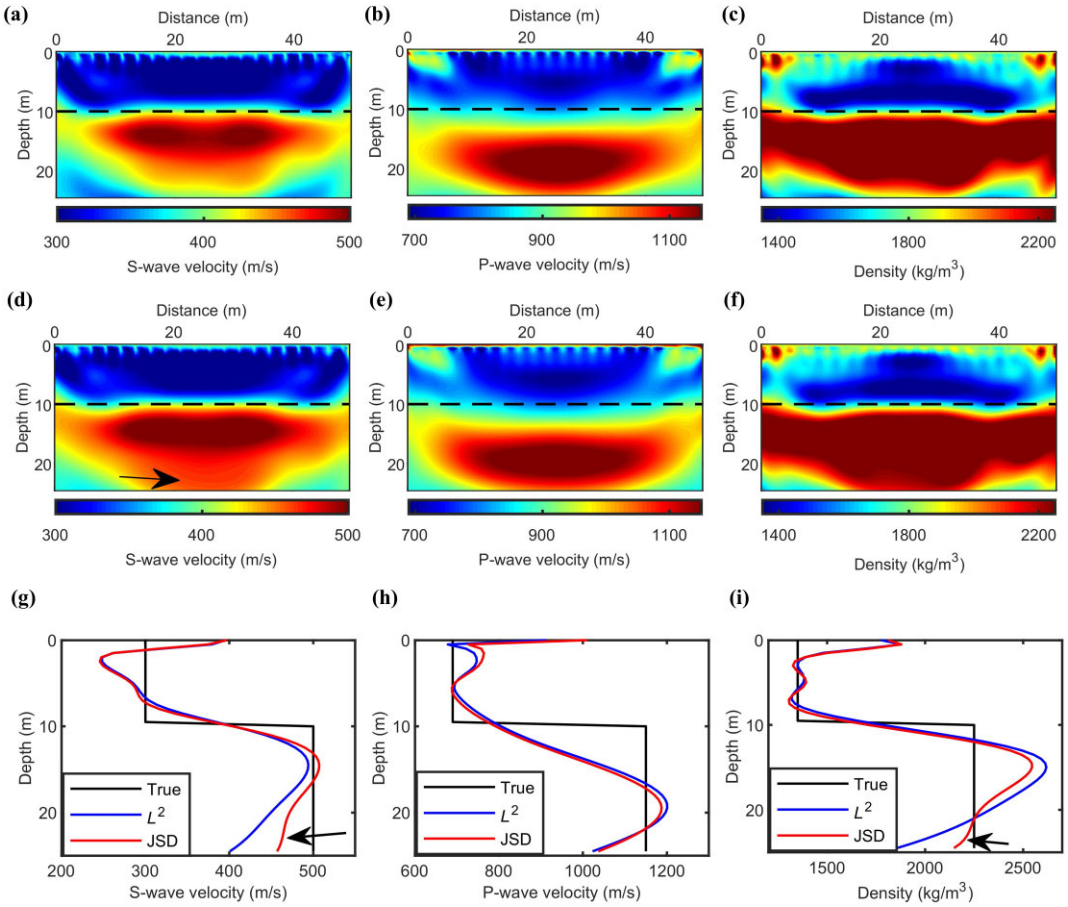

Figure 4. The inverted results of the reconstruction test for the two-layer model with noise-free data and poor initial models. The first row is the inverted results generated by the CWI, including the *S*-wave velocity (a), *P*-wave velocity (b) and density (c). (d), (e) and (f) denote the inverted results for V_S, V_P and ρ generated by our presented method, respectively. The black dashed lines in (a), (b), (c), (d), (e) and (f) draw the formation interface, and the black arrow in (d) points out the extra illumination enhancement part for *S*-wave velocity generated by our method. (g), (h) and (i) provide the 1-D comparisons of the inversion results (generated by the CWI and our presented method) and true model at the distance of 25 m for V_S, V_P and ρ, respectively. The true, inverted result generated by the CWI, and inverted result produced by our method at the distance of 25 m are represented by the black, blue and red curves in (g), (h), and (i), respectively. The black arrows in (g) and (i) indicate that more accurate values of *S*-wave velocity and density for deeper media can be attained by our method.

to compensate for the geometric diffusion difference between the predicted and observed data. In this case, the misfit is reformulated as:

$$E_{L^2}(p) = \frac{1}{2}\left\langle Lu^N(p) - d^N, Lu^N(p) - d^N \right\rangle, \tag{2}$$

where $u^{\mathrm{N}}(p)$ is the predicted data with normalization operation, and d^{N} is the normalized one for the observed data.

The JSD is from probability theory and statistics (Manning & Schütze 1999), which is a method of measuring the similarity between two probability distributions, defining by:

$$\mathrm{JSD}(P \| D) = \frac{1}{2}\left[P \ln\left(\frac{2P}{P+D}\right) + D \ln\left(\frac{2D}{P+D}\right)\right], \tag{3}$$

where P and D represent two probability distributions, and the above formula would reach the minimum value 0 when P and D overlap exactly. Since the seismic signal does not have the non-negative property, we cannot regard it as a probability distribution, directly, and a trace-by-trace normalization/standardization operation is still necessary. According to the different misfit definition for WI, the normalization operation in this paper is defined as:

$$s^N = \begin{cases} \frac{s}{\max(|s|)}, \text{ifmisfitisderivedby} L^2\text{norm}, \\ \frac{\frac{s}{\max(|s|)}+1.1}{\left\langle \frac{s}{\max(|s|)}+1.1 \right\rangle}, \text{ifmisfitisderivedbyJSD}. \end{cases} \tag{4}$$

In the above formula, s and s^{N} are the raw seismic signal and the normalized signal, respectively, 1.1 is an empirical parameter that maintains the non-negative property of the probability distribution of the seismic signal. The meaning of the symbol $<\ldots>$ is:

$$< s >= \int s(t)\mathrm{d}t. \tag{5}$$

For the misfit derived by the L^2 norm, the normalized operation would let each 1-D seismic signal in the inversion system to be normalized with respect to the corresponding maximum absolute amplitude. While for the misfit defined by the JSD, the normalized operation would make each 1-D seismic signal in the inversion system to be both non-negative and integrated to 1 along the time-axis. Similarly, we also give the misfit derived from the JSD for the observed data obtained from the single shot by the inner

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

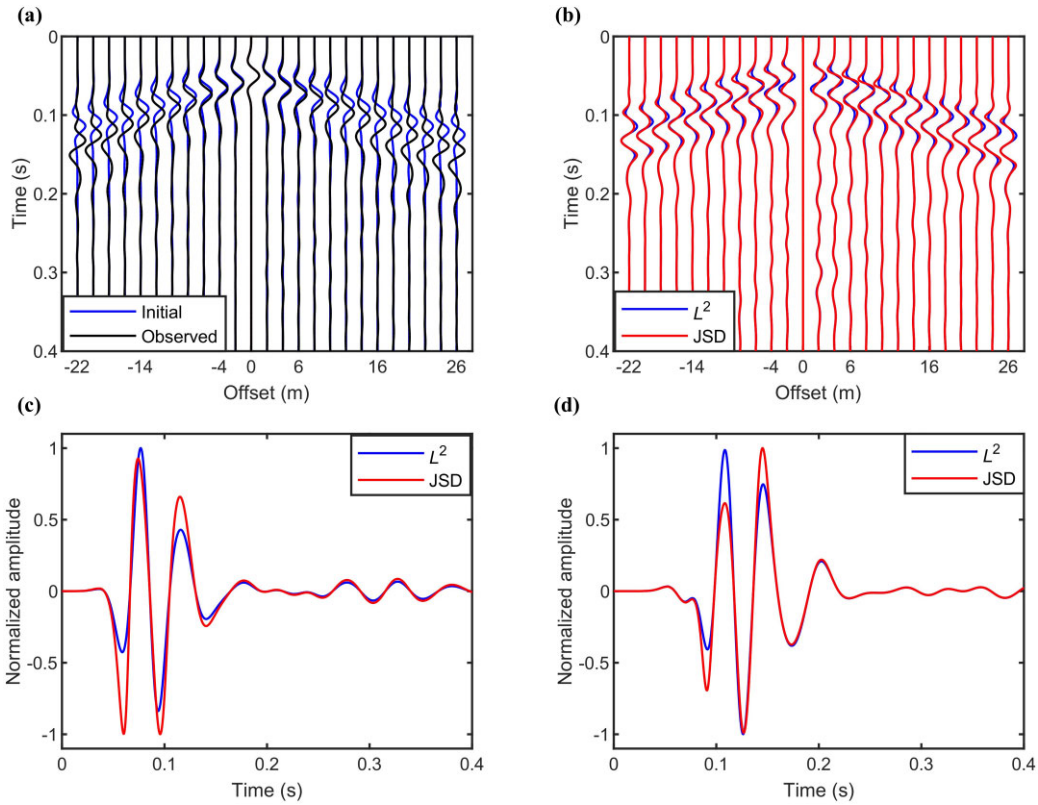

Figure 5. The initial and observed waveforms (a) of the 10th shot for the reconstruction of two-layer model with noise-free data and poor initial models. The adjoint-source waveforms of the 10th shot at the 0th iteration defined by the CWI and our method are shown in (b). For clarity, the trace-spacing and recording-time of (a) and (b) have been adjusted to 2 m and 0.4 s, respectively. Each waveform in (a) and (b) is normalized to 0.5 according its maximum absolute amplitude. (c) and (d) show the 1-D adjoint-source waveforms within 0.4 s at the offset of 10 and 20 m of the 10th shot at the 0th iteration for the CWI and our method, respectively. The blue and red curves in (b), (c) and (d) show the adjoint-source waveforms generated by the CWI and our method, respectively.

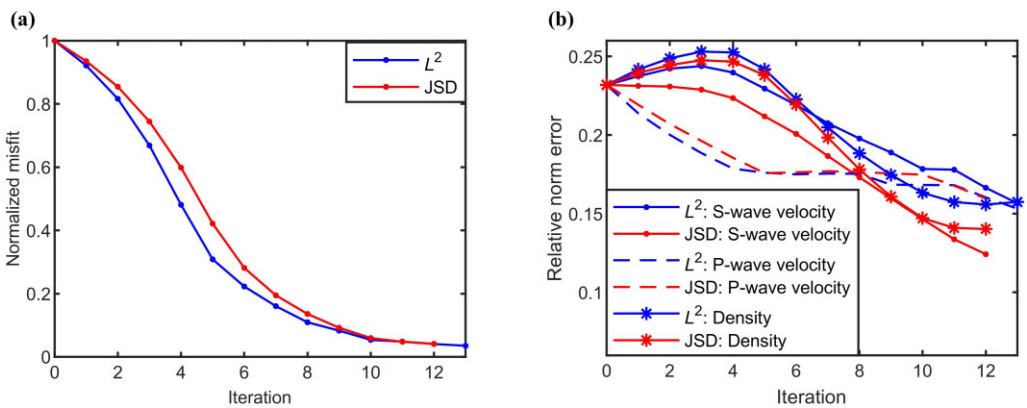

Figure 6. The descent curves (a) for the normalized misfit values against the iteration for the reconstruction of two-layer model with noise-free data and poor initial models, the blue line is the descent curve for the CWI, and the red line represents the descent curve for our presented method. Panel (b) shows the comparison of the descent curves of the relative norm error defined by our approach and the CWI during the entire iterative process. The blue and red curves in (b) show the corresponding results of the CWI and our method, respectively.

product, which is expressed as:

$$E_{\mathrm{JSD}}(p) = \frac{1}{2}\left\langle Lu^N(p), \ln\left(\frac{2Lu^N(p)}{Lu^N(p)+d^N}\right)\right\rangle + \frac{1}{2}\left\langle d^N, \ln\left(\frac{2d^N}{Lu^N(p)+d^N}\right)\right\rangle. \quad (6)$$

In particular, when there are multiple shot gathers in the inversion system, the total misfit function value is computed by superposing the misfit function value of each shot gather, simply.

2.2 Gradient and quasi-Hessian operator

FWI/WI is a large-scale optimization problem, limited by the attainable computing power, which is generally optimized by the local optimization algorithm (Métivier *et al.* 2014; Guan *et al.* 2022).

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

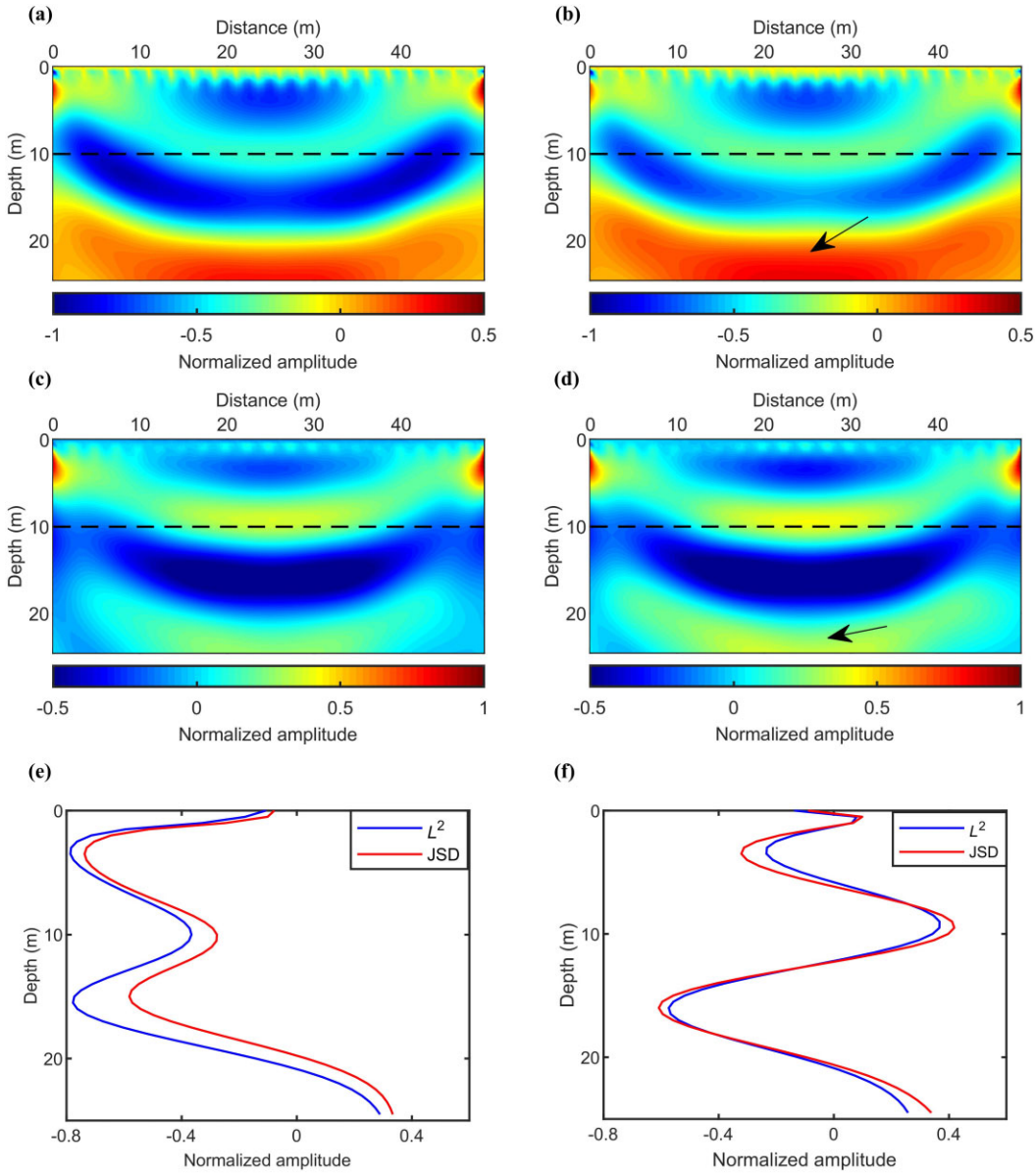

Figure 7. Comparison of the normalized descent direction at the 0th iteration between the CWI and the proposed MSWI for the reconstruction of two-layer model with noise-free data and poor initial models. Panels (a) and (c) are the normalized descent direction of the V_S and ρ of the CWI, (b) and (d) are the same display, but for our method. Panel (e) provides the 1-D comparison of the normalized descent direction of the *S*-wave velocity at the distance of 25 m between the CWI and proposed method, and (f) shows the same comparison result that belongs the density parameter. The blue and red curves in (e) and (f) show the corresponding results of the CWI and our method, respectively.

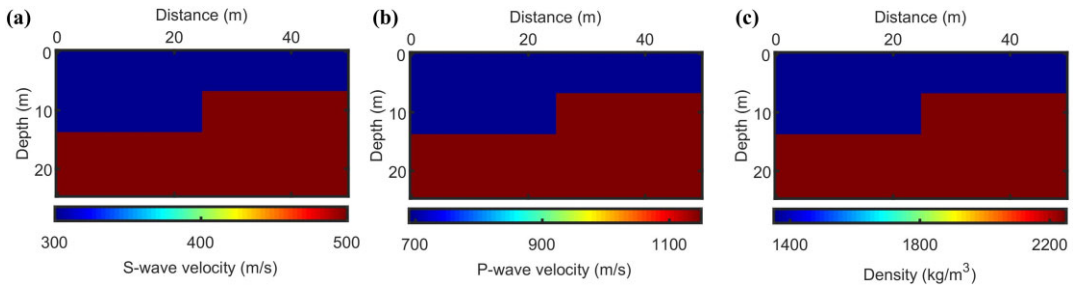

Figure 8. The true *S*-wave velocity (a), *P*-wave velocity (b) and density (c) of the reconstruction test for the fault model with noise-free data and poor initial models.

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

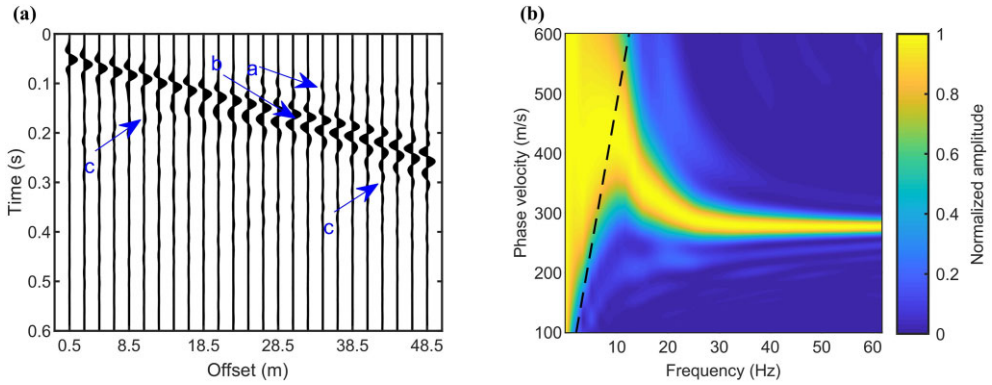

Figure 9. The normalized seismograms (a) for the first shot gather of the fault model reconstruction test with noise-free data and poor initial models, and the trace-spacing is adjusted as 2 m for clarity. The blue lowercases a, b and c with arrows in (a) indicate the refracted-wave, Rayleigh waves and scattered waves from the fault interface, respectively. Panel (b) is the dispersion spectrum of (a) calculated by the phase-shift method, and the black dashed line is defined by the length of the receiver-array of the first shot gather, which restricts the maximum measurable wavelength of the Rayleigh-wave.

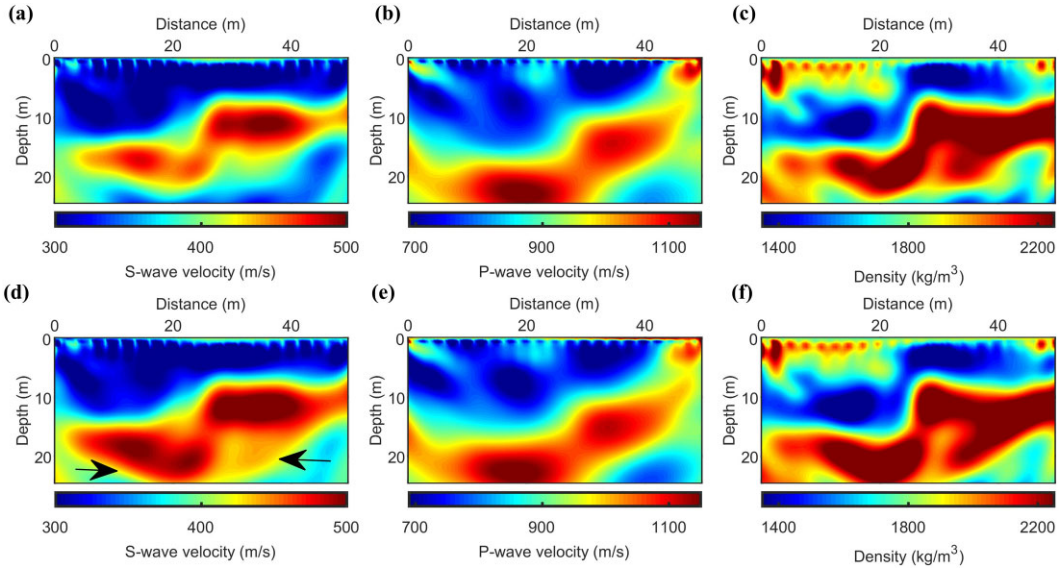

Figure 10. The inverted results of the reconstruction test for the fault model with noise-free data and poor initial models. The first row is the inverted results generated by the CWI, including the *S*-wave velocity (a), *P*-wave velocity (b) and density (c). Panels (d), (e) and (f) denote the inverted results for V_S, V_P and ρ generated by our presented method, respectively. The black arrows in (d) point out the extra illumination enhancement parts generated by our method.

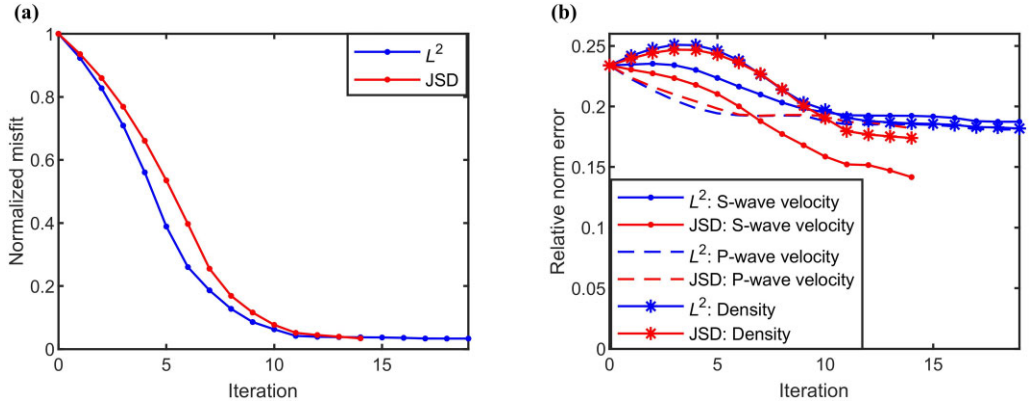

Figure 11. The descent curves (a) for the normalized misfit values against the iteration for the reconstruction of fault model with noise-free data and poor initial models, the blue line is the descent curve for the CWI, and that red line represents the descent curve for our presented method. Panel (b) shows the comparison of the descent curves of the relative norm error defined by our approach and the CWI during the entire iterative process. The blue and red curves in (b) show the corresponding results of the CWI and our method, respectively.

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

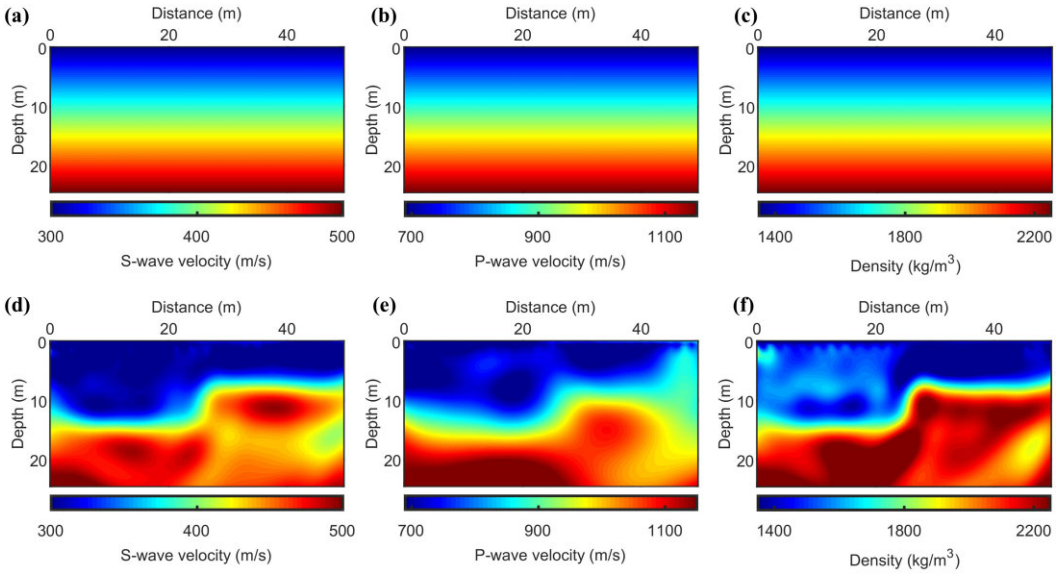

Figure 12. The more accurate initial V_S (a), V_P (b) and ρ (c) models for the reconstruction test of the fault model with noisy data, (d), (e) and (f) are the corresponding inverted results generated by our presented method.

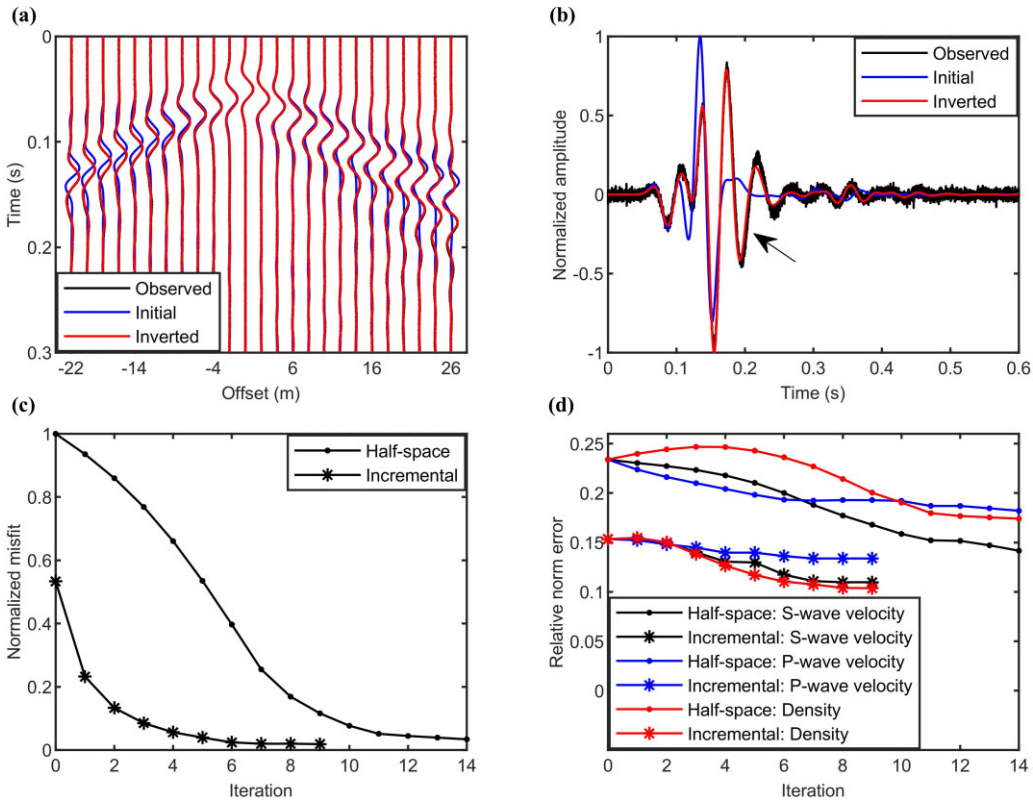

Figure 13. Exhibitions of the evaluation about the reconstruction test of the fault model with noisy data and more accurate initial models. The initial, inverted and observed waveforms (a) of the 10th shot, the trace-spacing and recording-time have been adjusted to 2 m and 0.3 s, respectively. Each waveform in (a) is normalized to 0.5 with respect to its maximum absolute amplitude. The single waveform fit (b) at the offset of 25 m of the 10th shot, and the black arrow indicates the scattered wave in the observed waveform. The descent curves (c) for the normalized misfit, the black dotted line is the result of the fault model reconstruction with noise-free data under the condition of the poor half-space initial models (the result in Section 3.1.2), the black star line represents the result of the fault model reconstruction with noisy data under the condition of the more accurate initial model settings, and its misfit values are normalized with respect to the initial misfit value of the black dotted line. The descent curves (d) for the relative norm error, including the result in Section 3.1.2 and this section. The string 'Half-space' in legend of (c) and (d) is the abbreviation for the half-space initial models, and the string 'Incremental' in legend of (c) and (d) is the abbreviation for the incremental initial model settings.

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

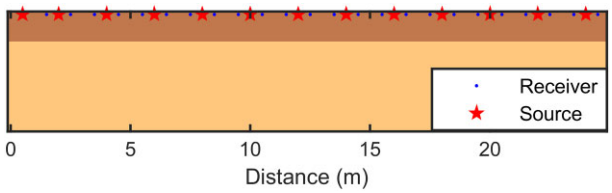

Figure 14. The schematic diagram of the observed system of the reconstruction test for the power tunnel model with noisy data and more accurate initial models, and the following field data case. Both the sources and receivers are configured on the free-surface, the red pentagram indicates the location of the source on the free-surface for each shot gather, and the blue dots represent the positions of the corresponding receivers.

The local optimization algorithms can be divided into two main categories: the gradient-based algorithms (it is usually called as the first-order optimization technique) and approximate Newtonian methods (utilizing the information of Hessian matrix and gradient of the misfit, and it is also named as the second-order optimization technique). The gradient vector (first-order partial derivatives) of the misfit of the single shot gather can be computed efficiently by the first-order adjoint-state method (Plessix 2006), which is as follows:

$$\frac{\partial E(p)}{\partial p} = -\left\langle \frac{\partial S(p)}{\partial p} u, \lambda^* \right\rangle, \tag{7}$$

where $S(p)$ is the partial-differential operator for the 2-D wave equation of P-SV type of elastic media in this paper, u is the forward wavefield, and λ^* represents the adjoint wavefield obtained by solving the adjoint wave equation. The detailed forms of the adjoint wave equation and derivation for the above formula are arranged at Appendix A.

According to studies, the inverse Hessian matrix plays an important role in balancing the gradient amplitude, illumination compensation, suppressing the cross-talk between different parameter classes and accelerating the convergence rate (Métivier *et al.* 2014). However, because of the huge dimension of the Hessian matrix, it can only be estimated by some implicit methods. Shin *et al.* (2001) estimated it in prestack depth migration by the inverse scattering theory, Operto *et al.* (2006) and Gao *et al.* (2023) implemented the Gauss Newton's estimation for it. Another strategy is to construct it through differentiating the gradient of the misfit from previous iterations, such as the *l*-BFGS algorithm (Nocedal & Wright 2006). To our knowledge, the most accurate computation scheme for the Hessian operator is the second-order adjoint state method (Métivier *et al.* 2014), while we only refer to the method of Shin *et al.* (2001) to simply estimate the quasi-version of the Hessian matrix of the misfit with the observed data obtained from the single shot, which is named as the quasi-Hessian operator in this paper, expressing as:

$$\frac{\partial^2 E(p)}{\partial p^2} \approx \left\langle \frac{\partial S(p)}{\partial p} u, \frac{\partial S(p)}{\partial p} u \right\rangle. \tag{8}$$

It can be noted that the quasi-Hessian operator in the above formula is only related to the forward wavefield u and partial-differential operator $S(p)$, therefore, the calculation of the quasi-Hessian operator is not affected by the definition of the misfit function. In other words, the quasi-Hessian operator in the CWI and our method is exactly the same, which means that the definition of the misfit function only affects the calculation of the gradient vector in the paper (eq. 7). More detailed derivation and interpretation about the above formula are also given in Appendix A. Similarly, the gradient vector and quasi-Hessian operator of the misfit function of the multiple shot data are still computed by stacking them of each shot gather, simply.

2.3 Preconditioned steepest-descent (PSD) algorithm

We refer to the core idea of the Levenberg–Marquardt (L–M) method of achieving the compromise between the steepest-descent direction and Gauss–Newton descent direction by introducing a hyper-parameter μ (Aster *et al.* 2018) to design the preconditioned steep-descent (PSD), which was originally applied to the inversion of the Rayleigh-wave dispersion curves (Yan *et al.* 2022b), and the particular iteration sequence for the 2-D MSWI in the paper is defined as:

$$p_{k+1} = p_k + \alpha_k d_k{}^N, \ k \geq 0, \tag{9a}$$

$$d_k{}^N = \left(d_k{}^S, d_k{}^P, d_k{}^D\right), \tag{9b}$$

$$d_k{}^S = -\frac{\min\left(V_{S_k}\right)}{\max\left(\left|\left(H_k{}^S + \mu_k \max\left|H_k{}^S\right|\right)^{-1} g_k{}^S\right|\right)} \left(H_k{}^S + \mu_k \max\left|H_k{}^S\right|\right)^{-1} g_k{}^S, \tag{9c}$$

$$d_k{}^P = -\frac{\min\left(V_{P_k}\right)}{\max\left(\left|\left(H_k{}^P + \mu_k \max\left|H_k{}^P\right|\right)^{-1} g_k{}^P\right|\right)} \left(H_k{}^P + \mu_k \max\left|H_k{}^P\right|\right)^{-1} g_k{}^P, \tag{9d}$$

$$d_k{}^D = -\frac{\min\left(\rho_k\right)}{\max\left(\left|\left(H_k{}^D + \mu_k \max\left|H_k{}^D\right|\right)^{-1} g_k{}^D\right|\right)} \left(H_k{}^D + \mu_k \max\left|H_k{}^D\right|\right)^{-1} g_k{}^D, \tag{9e}$$

where p_k represents the model parameter at the kth iteration, which consists of V_S, V_P and ρ. For the kth iteration, $d_k{}^{\mathrm{N}}$ is the amplitude-adjustive descent direction, which includes $d_k{}^{\mathrm{S}}$, $d_k{}^{\mathrm{P}}$ and $d_k{}^{\mathrm{D}}$, $d_k{}^{\mathrm{S}}$, $d_k{}^{\mathrm{P}}$ and $d_k{}^{\mathrm{D}}$ are the descent directions for V_S, V_P and ρ after the scaled operation according to the corresponding minimum, respectively; $g_k{}^{\mathrm{S}}$, $g_k{}^{\mathrm{P}}$ and $g_k{}^{\mathrm{D}}$ represent the corresponding gradient vector of the misfit function about V_S, V_P and ρ, respectively; $H_k{}^{\mathrm{S}}$, $H_k{}^{\mathrm{P}}$ and $H_k{}^{\mathrm{D}}$ denote the corresponding quasi-Hessian operator of the misfit function about V_S, V_P and ρ, respectively; here, we only use the diagonal elements of the quasi-Hessian operator for getting the inverse matrix, easily. Certainly, Pratt *et al.* (1998) also pointed out the Hessian matrix is diagonal dominant for high-frequency approximation. μ_k is the hyper-parameter that controls the compromise degree between the steepest-descent direction and quasi-Newton descent direction at the kth iteration, when the selected $\mu_k$approaches to zero, the descent direction is approximately the quasi-Newton descent direction, and the descent direction would be the steepest-descent direction when μ_k is close to positive infinity. The significant hyperparameter would be determined by trials at each iteration, and ranging from 10^{-6} to 10 in the paper. α_k is the inexact percentage step-length (expressing as a percentage of the respective minimum values of the current V_S, V_P and ρ model) that is chosen by the Armijo search (Nocedal & Wright 2006) at the kth iteration.

To sum up, we can give the specific algorithm implementation scheme of the PSD to optimize the 2-D MSWI problem in the paper, which is as follows:

Algorithm 1: Preconditioned steepest-descent (PSD) algorithm

Step 1: Let $k = 0$, given the initial model p_k that consists of the initial S-wave velocity, P-wave velocity and density model, calculating the misfit function value E_k of p_k, and setting the maximum number *maxIter* of iterations.
Step 2: If $k > maxIter$, the iteration would be terminated, and returning the inverted result p_k.
Step 3: Calculating the gradient vector g_k and quasi-Hessian operator H_k of the misfit function about the S-wave velocity,

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

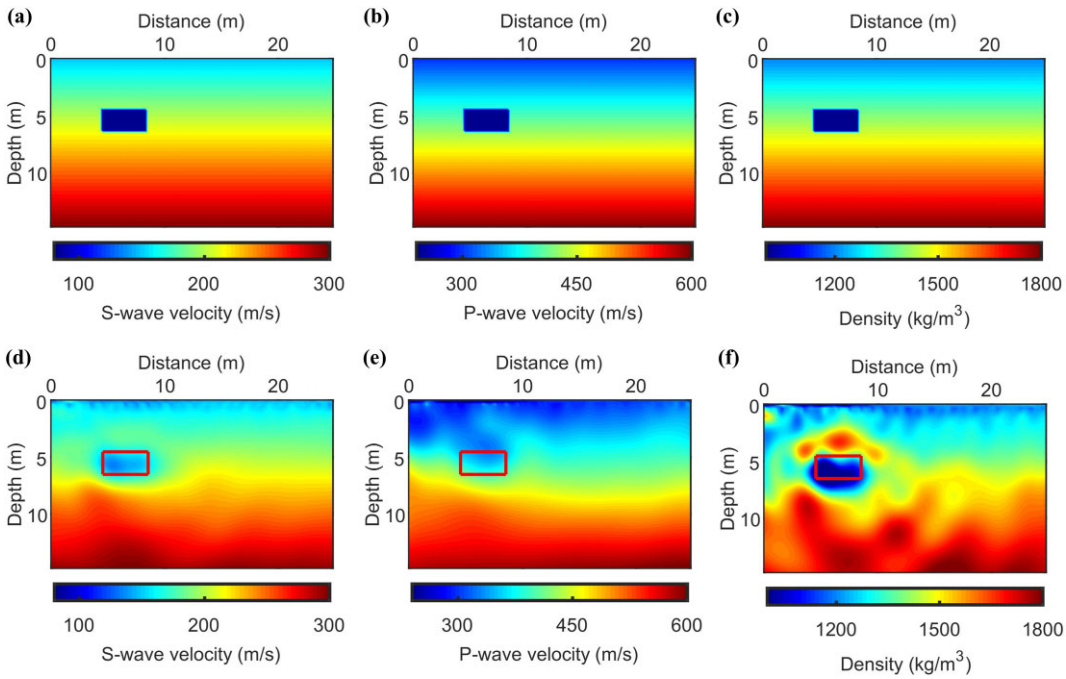

Figure 15. The true *S*-wave velocity (a), *P*-wave velocity (b) and density (c) of the reconstruction test for the power tunnel model with noisy data and more accurate initial models, (d), (e) and (f) are the corresponding inverted results generated by our method, and the red rectangular lines delineate the position of the power tunnel.

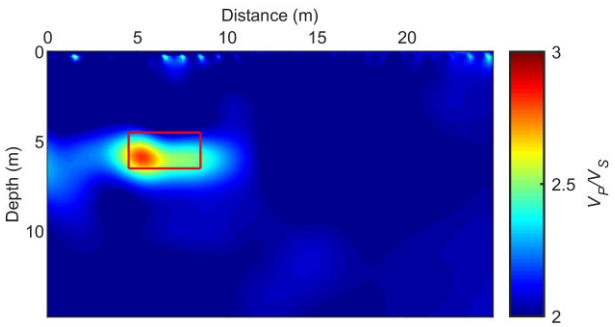

Figure 16. The inversion image of the ratios between the *P*-wave velocity and *S*-wave velocity for the reconstruction test of the power tunnel model with noisy data and more accurate initial models, which is generated by our presented method.

P-wave velocity and density according to the formula (7) and (8).

(a) Outer loop about the hyper-parameter μ_k: Given the trial sequence for μ_k, it is set as (10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1.1, 10) in the paper. For each trial μ_k, an alternative descent direction d_k^{N} would be generated.

(b) Inner loop about the percentage step-length α_k: Given the trial sequence for α_k, and the maximum trial value does not exceed 0.1. To perturb the current model p_k with the trial α_k and the trial descent direction d_k^{N}, then, calculating and saving the misfit function values for these trial models in this loop.

Step 4: To find the candidate model $p_{\min}$ from these trial models occurred in Step 3 that has the smallest misfit function value $E_{\min}$, if $E_{min} < E_k$, $k = k + 1$, $p_k = p_{\min}$, $E_k = E_{\min}$, for field data, the source wavelets of each shot gather are also updated by the source correction filter (Groos *et al.* 2014), then go back to Step 2 and 3; otherwise, terminating the iteration and returning the inverted result p_k.

Please note that the predicted wavefield u is a function of the model parameter and source wavelet, thus the misfit function for the 2-D MSWI is also related to these two terms. Therefore, for field data case, the source wavelets of each shot gather would also be updated at each iteration, dynamically.

3 SYNTHETIC TESTS

3.1 Model reconstruction tests with noise-free data and poor initial models

To quantitatively evaluate the performance of our presented 2-D MSWI and the CWI, we first consider the reconstruction tests for the two-layer model and fault model with noise-free data and poor initial models (half-space models) in this part. Based on the structural similarity between different parameter classes, the true V_S, V_P and ρ model would share the same structure in the synthetic tests in the paper. The calculation domain for these two models has a grid number of 100 along the *x*-direction and 50 along the *z*-direction with the spatial-grid length at 0.5 m, and the schematic diagram of the observed system for these two synthesized tests is shown by Fig. 1. Both the sources and receivers are configured on the free-surface, and the number of the sources is 20 with the source-spacing at 2.5 m starting from 1.0 m along the *x*-direction. For each shot gather, there are 100 Z-component records with the total recording time of 0.6 s and sampling rate at 10 000 Hz. A 20 Hz Ricker wavelet with the delaying time of 50 ms is loaded as a source on the *Z*-component for each shot gather, and the source wavelets are considered as known terms for the synthetic tests in the paper. To assess the closeness of the inversion result to the true model, we introduce the relative norm error err_p, which is defined as:

$$err_p = \frac{\|p_{\mathrm{true}} - p_{\mathrm{inv}}\|_2}{\|p_{\mathrm{true}}\|_2}, \tag{10}$$

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

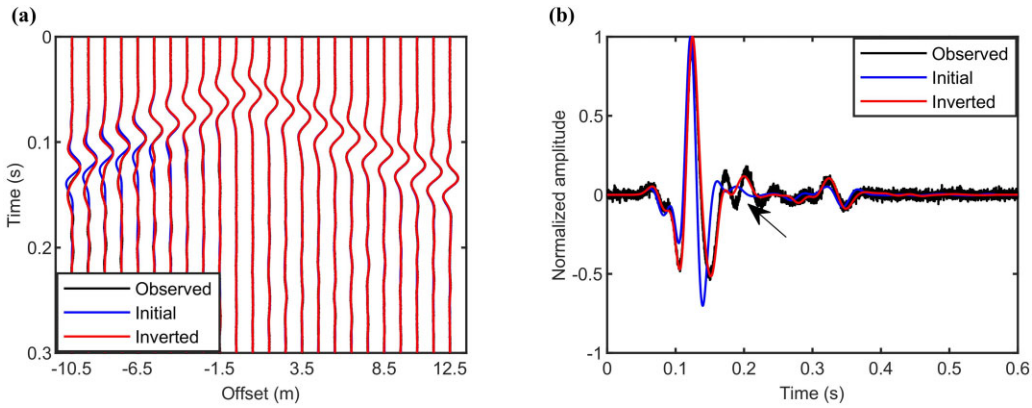

Figure 17. Exhibitions of the evaluation about the reconstruction test of the power tunnel model with noisy data and more accurate initial models. The initial, inverted and observed waveforms (a) of the 7th shot, and the recording-time has been adjusted to 0.3 s for clarity. Each waveform in (a) is normalized to 0.5 with respect to its maximum absolute amplitude. The single waveform fit (b) at the offset of −10.5 m of the 7th shot, and the black arrow indicates the scattered wave in the observed waveform.

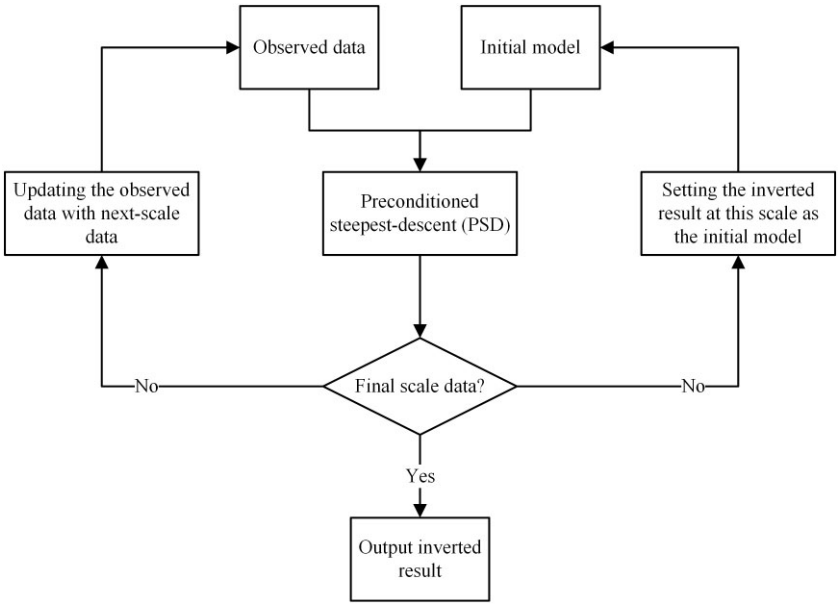

Figure 18. The detailed analysis workflow of the 2-D MSWI based on the JSD for field data.

where p_{true} and p_{inv} denote the true model and inverted result, respectively.

3.1.1 Reconstruction test for the two-layer model with noise-free data and poor initial models

The true and initial models of the reconstruction test for the two-layer model are given by Fig. 2. The V_S, V_P and ρ of the two-layer model have the same structure (Figs 2a, b and c), and the interface depth is set as 10 m. The first-layer and second-layer *S*-wave velocity are 300 and 500 m s^{-1} (Fig. 2a), respectively. The ratio between V_P and V_S is 2.3 (Fig. 2b), and the densities for the first-layer and second-layer are 1350 and 2250 kg m^{-3} (Fig. 2c), respectively. To better reveal the illumination ability to the deeper-media parameters, the initial models are considered as the half-space models, which are the *S*-wave velocity of 400 m s^{-1}, *P*-wave velocity of 920 m s^{-1} and density of 1800 kg m^{-3} (Figs 2d, e and f). Fig. 3 shows the synthetic records of the first shot gather and the corresponding dispersion spectrum calculated by the phase-shift method (Park *et al.* 1998), the Rayleigh-wave events are dominant, the refracted and scattered waves are faint, but visible (Fig. 3a). The black dotted lines of Fig. 3(b) are the analytic phase velocities calculated by the multimode fast vector transfer method (Fan *et al.* 2007), the fundamental-mode Rayleigh-wave is dominant for the frequency-band of 10–60 Hz, furthermore, the dispersion energy and the analytic phase velocities of the fundamental-mode Rayleigh-wave almost completely match, which verifies the reliability of our designed simulation program for the 2-D wave equation of *P*-SV type in elastic media. Here, it is necessary to give the empirical law of the energy distribution of the surface-wave dispersion spectrum. According to our knowledge, the energy distribution of the dispersion spectrum is mainly related to the location of sources and the subsurface structures. The subsurface sources would enhance the energy of the higher-mode surface waves in the dispersion spectrum with the condition of the same structure. When the source is deployed on the free-surface, the fundamental-mode surface wave is dominant for the incremental velocity structures, therefore, we only see the dispersion energy of the fundamental-mode surface-wave in

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

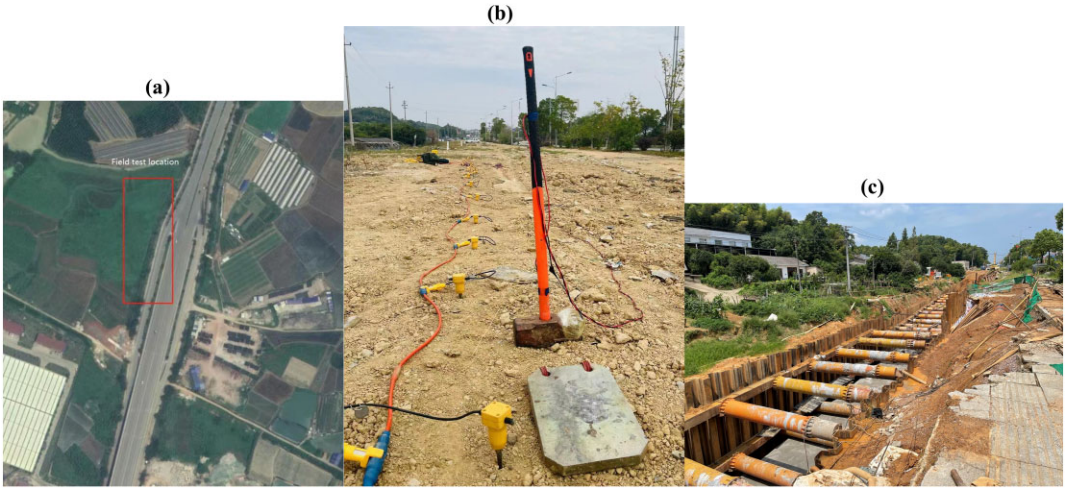

Figure 19. The detailed exhibitions about the survey area, and the observed system settings are the same as that of the above reconstruction test of the power tunnel model (Fig. 14). Panel (a) is the satellite map for the measurement area, and the red rectangle describes the approximate area for the survey line. Panel (b) is the photo taken during the data collection, and the 4.5 Hz *Z*-component receivers are chosen to record the seismic waves excited by a 10-kg sledgehammer, vertically. Panel (c) is the photo taken when the power tunnel was installed.

Fig. 3(b). For the structures that contain the high/low-velocity interlayers, the energy of higher-mode surface waves would be reflected in the dispersion spectrum when the source is still deployed on the free-surface.

The inversion results of the reconstruction test for the two-layer model are given through iteration, as shown by Fig. 4, where the images at the first row are the inverted results generated by the CWI (Figs 4a, b and c), and the images at the second row are the inverted results given by our proposed method (Figs 4d, e and f). The stratigraphic interface at the depth of 10 m can be accurately identified from the images of the inverted results of *S*-wave velocity and density (Figs 4a, c, d and f), respectively, while the correct depth of the interface is not clearly revealed from the inverted images of *P*-wave velocity (Figs 4b and e). The inaccurate *P*-wave velocity reconstruction results are mainly caused by the weak events associated with *P*-wave velocity in the observed waveforms, which also indicates the lower sensitivity of 2-D MSWI system to the V_P model. Compared with the CWI, our proposed method implements an extra illumination for the reconstruction of *S*-wave velocity and density (Figs 4a, d, g and i), and the black arrows in Figs 4(d), (g) and (i) point out the specific additional illumination parts. Specifically, our method enhances the ability to reveal the accurate values of *S*-wave velocity and density for deeper media (Figs 4d, g and i). For the *P*-wave velocity reconstruction, there are no significant differences between the CWI and our presented method (Fig. 4h). By comparing the initial and observed waveforms of the 10th shot gather, the discrepancy between the two waveforms is more obvious as the offset becomes larger, even exceeding half period (Fig. 5a). Fig. 5(b) gives the comparison for the adjoint-source waveforms of the 10th shot at the 0th iteration defined by our method and the CWI, and the adjoint-source waveforms of the two methods have similar traveltime characteristics, but the peaks and valleys that do not coincide along the time-axis. To compare the difference in detail, the 1-D adjoint-source waveforms at the offset of 10 and 20 m of the 10th shot are shown (Figs 5c and d), our proposed method achieves an implicit weighting along the time-axis for the 1-D adjoint-source waveforms of the CWI, including a strengthening for the trailing wave in the main event (Figs 5c and d). The enhancement for the trailing wave of the adjoint sources is mainly due to the nice measurement ability of the JSD on the detail modes of the signal, which would correspond to the different Fourier expansion coefficient in the frequency domain compared with that of the CWI. The different time–frequency characteristics for these two adjoint sources would be amplified by the adjoint wave equation, resulting in different adjoint wavefields and thus different gradients. Moreover, the difference between the 1-D adjoint-source waveforms defined by the two methods also becomes more significant as the offset becomes larger (Figs 5c and d).

To quantitatively compare the performance of our method with that of the CWI, the descent curves for the misfit values and the relative norm error of the entire iteration are shown (Figs 6a and b). We know that the parameters of the deeper medium contribute less to the misfit function value compared with the shallow-media parameters for the observed system deployed on the free-surface, and the intensity of the update for deeper-media parameters is enhanced for our method, therefore, the convergence rate of the misfit function of our method is slightly lower than that of the CWI (Fig. 6a). Nonetheless, both our approach and the CWI reach about 0.05 of the initial misfit value through iteration (Fig. 6a), but a better reconstruction result is generated by our method (Fig. 4). For the reconstruction of *S*-wave velocity, the performance of our proposed method is approximately 1.3 times that of the CWI from the perspective of closing to the true model (Fig. 6b). For the density reconstruction, the performance of our presented method is about 1.2 times that of the CWI (Fig. 6b). For the *P*-wave velocity reconstruction, the performance of the two methods is almost equivalent (Fig. 6b). The computational efficiency is also an important indicator for evaluating the performance of a novel method, and the cost of computing resources are not increased for our method under the condition of improving the imaging accuracy. Specifically, the computation time of the reconstruction test of the two-layer model by the CWI is 4650.4 s, and that of our method is 4300.6 s with the same hardware environment. We can also give such inference that the inversion system has almost the strongest sensitivities for *S*-wave velocity and density, while the poorest sensitivity to the *P*-wave velocity (Fig. 6b).

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

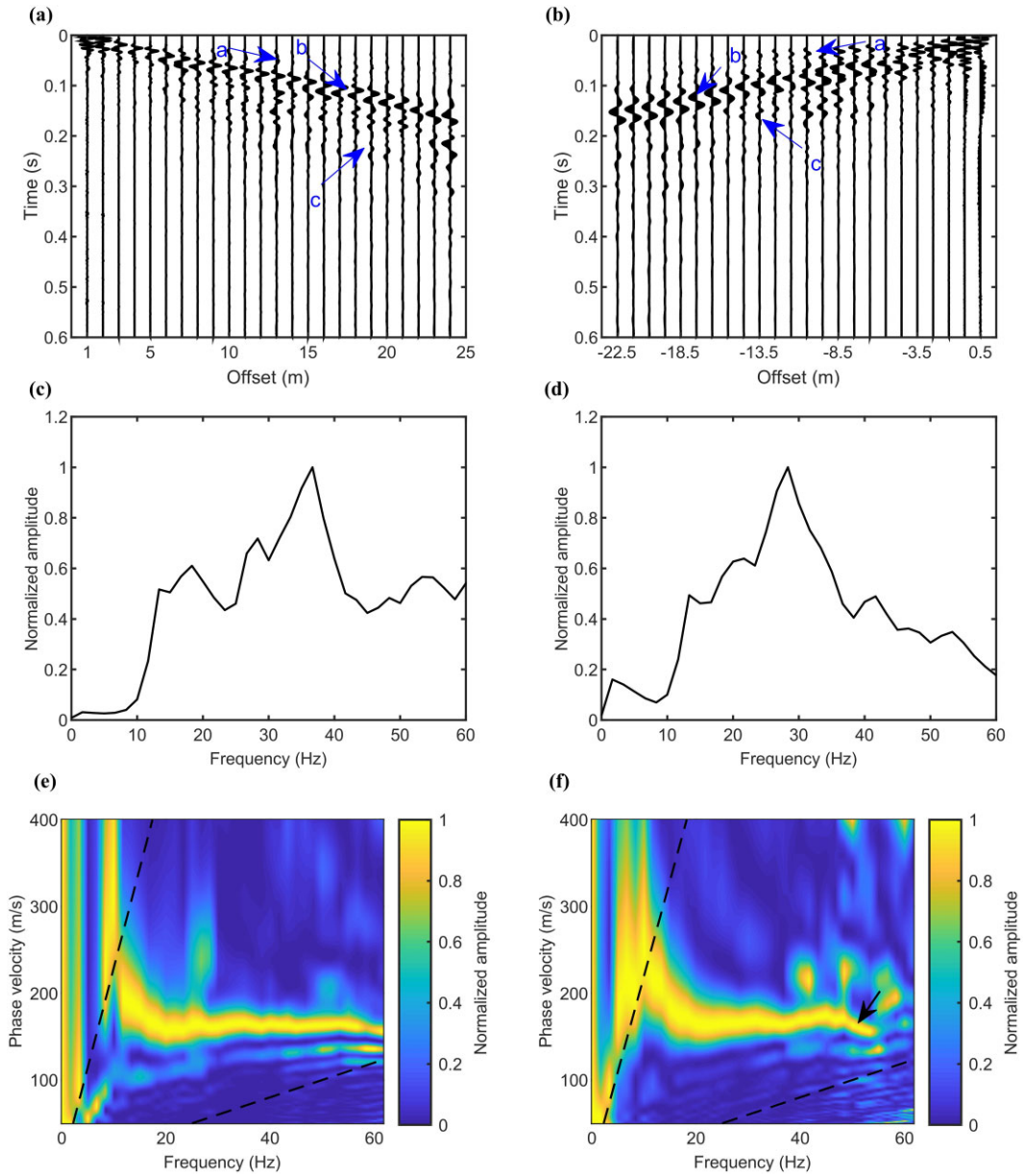

Figure 20. The characteristic displays for the field seismic-wave data. Panels (a) and (b) are the normalized seismograms of the first (the source is excited at left) and 13th shot (the source is excited at right) gather, respectively. The blue lowercases a, b and c with arrows in (a) and (b) indicate the refracted-wave, Rayleigh waves and scattered waves. Panels (c) and (d) are the corresponding normalized amplitude spectrum in the frequency-domain for the records in (a) and (b), respectively. The dispersion spectra (e) and (f) are extracted from the records in (a) and (b), respectively, and the black arrow indicates the phase velocity discrepancies at the high-frequency band. The two black dashed lines in (e) and (f) restrict a reliable area for picking the phase velocities with the Nyquist wavenumber (0.5/dx, dx is the receiver-spacing) and maximum wavelength (equals to the array length).

To explore the source of the better inversion results of the V_S and ρ generated by our method, the normalized descent directions for them at the 0th iteration are shown by Fig. 7. Compared with the normalized descent direction of the *S*-wave velocity defined by the CWI, that derived by our method has the wider area with large positive-amplitude when the depth is greater than 15.0 m (Figs 7a and b), as shown by the black arrow in Fig. 7(b), which is more in line with our expectations. The area of the bad distribution with the negative-amplitude in Fig. 7(b) at the approximate depth range of 10.0–15.0 m is smaller than that of Fig. 7(a). Although, the area with negative-amplitude above the depth of 10.0 m in Fig. 7(b) is slightly smaller than that in Fig. 7(a), our method provides a better descent direction for the V_S, entirely. The similar features are also reflected in the normalized descent direction of the density with slightly weaker degree (Figs 7a, b, c and d), and the black arrow in Fig. 7(d) also points out the wider area with large positive-amplitude. The 1-D comparisons for the normalized descent direction of the *S*-wave velocity shows that our method defines a better amplitude range below the depth of 10 m, and the amplitude range above the depth of 10 m is very similar to the CWI (Fig. 7e). The same characteristic is also identified in the 1-D comparisons for the normalized descent direction of the density, and

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

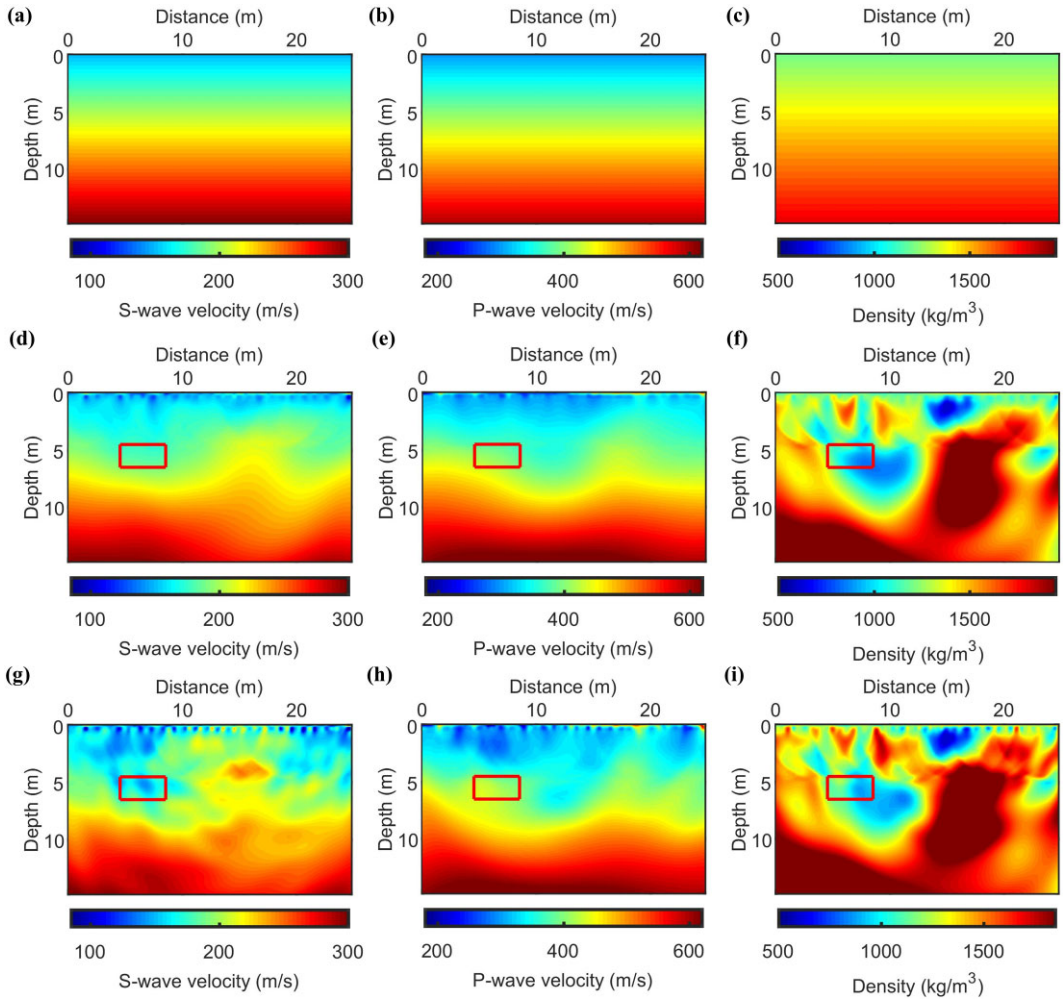

Figure 21. The initial models and inverted results of the field data generated by our presented 2-D MSWI based on the JSD. The initial *S*-wave velocity (a) is constructed by the dispersion spectrum analysis, (b) and (c) are the initial *P*-wave velocity and density, respectively. Panels (d), (e) and (f) are the inverted V_S, V_P and ρ at the first scale of 5 to 20 Hz, respectively. Panels (g), (h) and (i) show the inversion images of V_S, V_P and ρ at the second scale of 5–50 Hz, respectively. The red rectangles in (d), (e), (f), (g), (h) and (i) denote the burial position for the power tunnel given by the excavation result.

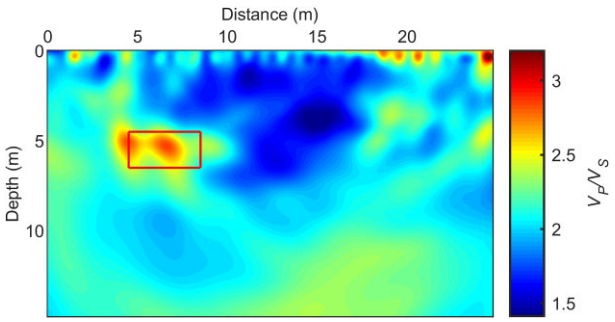

Figure 22. The inverted V_P/V_S profile of the field data generated by our presented 2-D MSWI based on the JSD. The red rectangle represents the burial position given by the excavation result.

the degree of this feature is relatively weak (Fig. 7f). In short, our method provides the better descent direction for the S-wave velocity and density, and eventually generating the better inversion results for them. In essence, the improvement benefits from the nice measurement for the detail modes of the seismic signal by the JSD, which is manifested as the enhanced trailing wave for the corresponding 1-D adjoint-source waveforms (Figs 5c and d). As a result, our method realizes the weighting for the 1-D adjoint-source waveforms of the CWI along the time-axis, then reforming the time–frequency characteristics of the adjoint wavefield and thus generating the different gradient vectors.

3.1.2 Reconstruction test for the fault model with noise-free data and poor initial models

Based on the structural similarity between different parameter classes, the true *S*-wave velocity, *P*-wave velocity and density of the reconstruction test for the fault model with noise-free data and poor initial models are also set to have the same structure (Fig. 8). The initial models are also considered as the half-space models, and the detail settings for the initial *S*-wave velocity, *P*-wave velocity and density remain the same as that of the above reconstruction test for the two-layer model (Figs 2d, e and f). Similarly, the Rayleigh-wave events still dominate in the synthetic waveforms for the first shot gather, the energy of refracted waves is still weak, in particular, the scattered waves become more abundant because of the existence of

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

Table 1. The detailed description of the borehole data and *S*-wave velocity range for each formation near the survey area.

Serial no.	From (m)	To (m)	Formation and description	*S*-wave velocity (m s^{-1})
1	0.0	3.0–7.0 m	The miscellaneous fill layer composed of loose deposits	100–200
2	3.0–7.0 m	8.0 m	The silty clay layer interspersed with miscellaneous fill	100–300
3	8.0 m	17.0 m	The sedimentary layer composed of strong weathering slate	200–300

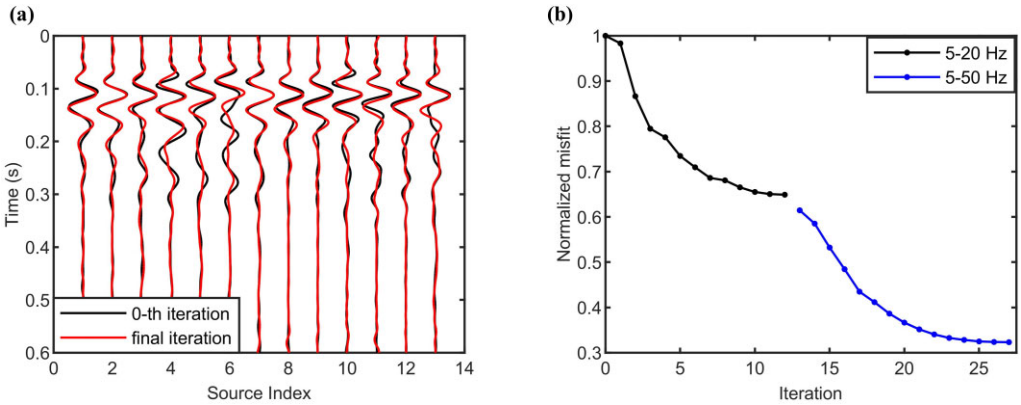

Figure 23. The source estimations (a) at the 0th iteration of the first scale, which are calculated by the source correction filter. For clarity, each source wavelet in (a) is normalized to 0.5 with respect to its maximum absolute amplitude. The descent curves (b) of the normalized misfit values for the first scale and second scale, and the misfit values are normalized with respect to the initial value.

fault interface (Fig. 9a). However, only the dispersion energy for the fundamental-mode Rayleigh-wave is reflected in the dispersion spectrum generated by the phase-shift method, while there are almost no additional responses about the scattered waves (Fig. 9b), which also shows that the classical dispersion inversion based on the assumption of 1-D structure would be difficult to reveal the fault interface.

Fig. 10 shows the inverted results generated by the CWI and our approach, and the correct location of fault interface can be inferred from the inversion images of *S*-wave velocity and density (Figs 10a, c, d and f), but it cannot be revealed by the inverted *P*-wave velocity (Figs 10b and e). Compared with the CWI, our method realizes the illumination enhancement for *S*-wave velocity, again (Figs 10a and d), as indicated by the black arrows (Fig. 10d), furthermore, the better descriptions for the right part of fault model are also achieved by our method (Fig. 10d). Similarly, because the update for the deeper medium parameters of our method is enhanced, the convergence rate of the misfit of our method is also slightly lower than that of the CWI (Fig. 11a). These two methods can reach almost the same misfit value through iteration (Fig. 11a), but the revealed inversion results are obviously different (Fig. 10), which illustrates that the path for the model update is obviously different due to the nature of the corresponding method. For the inversion results of *S*-wave velocity, the parameter reconstruction accuracy of our method is about 1.3 times that of the CWI, and the reconstruction result of density is also slightly improved (Fig. 11b). While for the reconstruction of *P*-wave velocity, both methods have almost the same performance (Fig. 11b). The computational cost of the both methods is still roughly consistent. Specifically, the computation time of the fault model reconstruction by the CWI is 4194.1 s, and that of our method is 4069.1 s. According to the descent curves of the relative norm error during the entire iteration process (Fig. 11b), the conclusion that the inversion system has strongest sensitivities to the *S*-wave velocity and density, but the poorest sensitivity to the *P*-wave velocity can also be obtained, again.

3.2 Model reconstruction tests with noisy data and more accurate initial models

The above reconstruction tests for the two-layer model and fault model with noise-free data and poor initial models have verified that the proposed method is better than the classical method under the same conditions. However, for more practical situations, the initial guess is typically not set as the poor half-space, but more accurate initial models can be built through the dispersion spectrum analysis for Rayleigh waves, and the observed waveforms would contain noise, inevitably. In view of this, we design the reconstruction test for the fault model with noisy data under the condition of the better initial models in the following part, and other settings remain the same as the above reconstruction test for the fault model with noise-free data and poor initial models. Also, we conduct a reverse test for the field case of power tunnel detection, and the parameter settings related to *S*-wave velocity, *P*-wave velocity and density are derived from the inverted results of field data. Similarly, the better initial model settings can also be implemented for the reconstruction of power tunnel model through the dispersion spectrum analysis. The anti-noise capability of our proposed method would be revealed by the two synthetic tests. The noise in the data of these two synthesized tests is considered as the white Gaussian noise with the signal-to-noise ratio at 20.

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

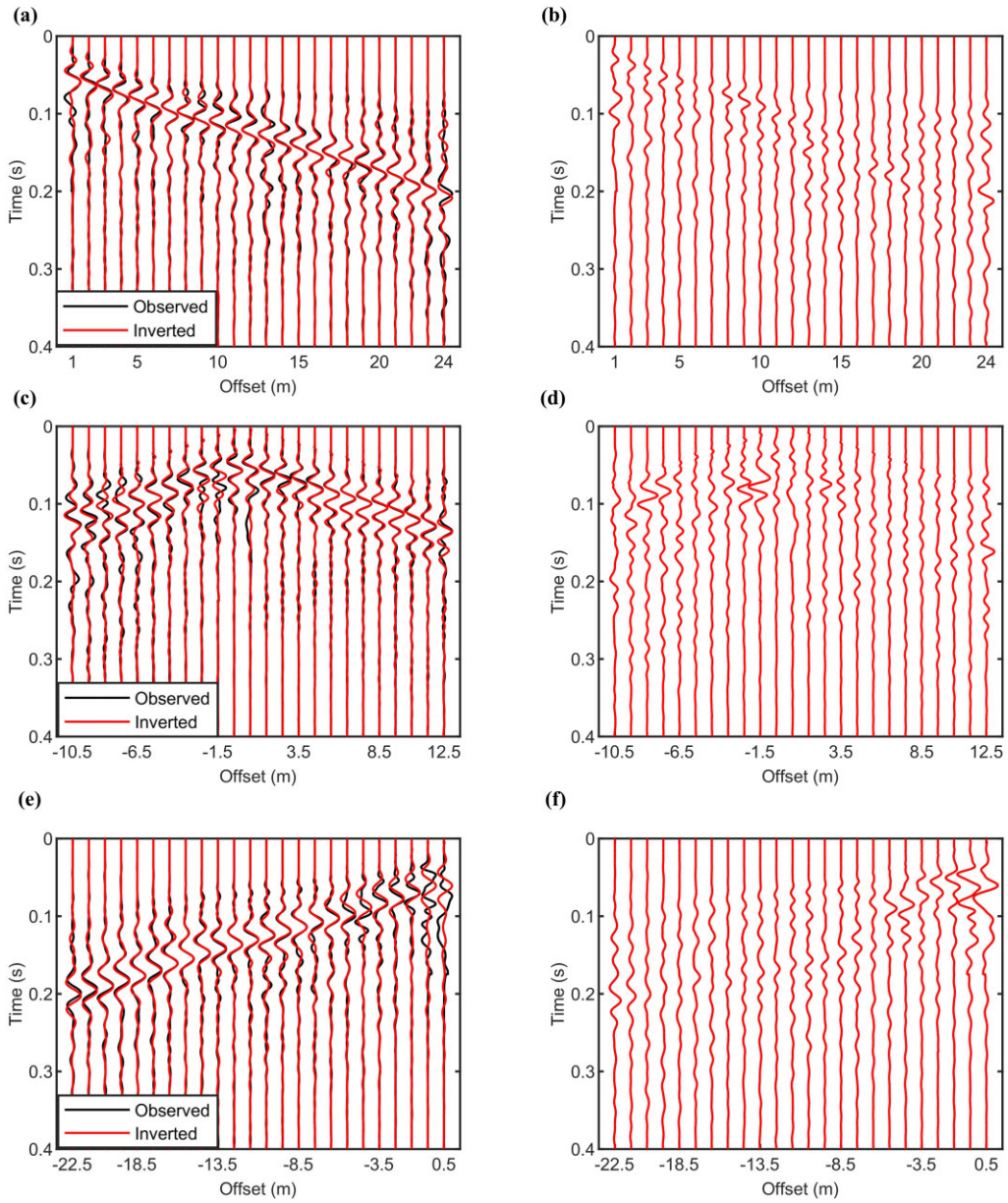

Figure 24. Comparisons between the observed and the corresponding inversion waveforms at the second scale of the field data inversion. Panel (a) gives the normalized waveform comparison between the observed waveforms at the second scale of the first shot and the corresponding inversion waveforms. Panel (b) shows the residuals between the observed and inversion waveforms in (a). Panels (c) and (d), and (e) and (f) are the same exhibitions as (a) and (b), but for the 7th and 13th shot gather. For clarity, the waveforms in (a), (c) and (e) are normalized to 0.5 with respect to the corresponding maximum absolute amplitude.

3.2.1 Reconstruction test for the fault model with noisy data and more accurate initial models

The better initial models for the reconstruction test of the fault model may be built through the dispersion spectrum analysis of Rayleigh waves, here, we construct the 1-D linear incremental models as the initial guess (Figs 12a, b and c). Although the observed waveforms contain noise, the inverted results for the *S*-wave velocity, *P*-wave velocity and density generated by our proposed method have achieved a fine description for the fault interface due to the better initial model settings (Figs 12d, e and f). Moreover, the more accurate description for deeper media has also been implemented, compared with the test above (Figs 10d, e and f; Figs 12d, e and f). The difference between the initial and observed waveforms of the 10th shot is significantly less than half period, and the inverted waveforms have no difference from the observed waveforms, visually (Fig. 13a). The single waveform fit at the offset of 25 m of the 10th shot shows that the trailing wave or scattered wave in the main event contains the abundant information related to the parameter distribution of underground medium (Fig. 13b), specifically, the fault interface is mainly reflected by the scattered waves for this reconstruction test. Compared with the corresponding test with the poor half-space initial models, the final misfit value and relative norm error become smaller due to the better initial model settings (Figs 13c and d). The conclusion that the inversion system has the strongest sensitivities to the *S*-wave velocity and density, and the poorest sensitivity to the *P*-wave velocity can still be confirmed (Fig. 13d), again.

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

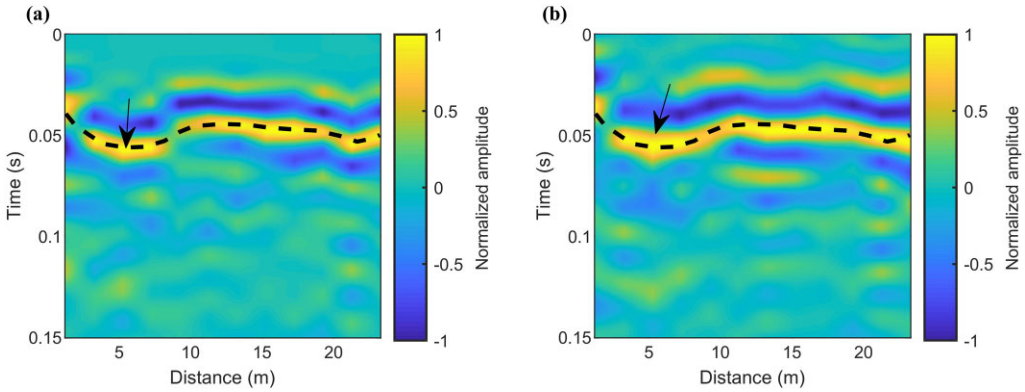

Figure 25. Common-offset profile comparisons between the predicted data of the inversion results and observed data at the second scale. (a) The common-offset profile of the observed data of the second scale. (b) The common-offset profile of the predicted data of the inversion results at the second scale. The common-offset parameter of (a) and (b) is 2.5 m, and each common-offset channel is placed at the midpoint position between the corresponding source and channel. The black dashed line in (a) is the traveltime curve at the common-offset 2.5 m picked up according to the maximum amplitude principle. The traveltime curve at the common-offset 2.5 m of the observed data is drawn in the common-offset profile of the predicted data, as shown by the black dashed line in (b). The black arrows in (a) and (b) indicate the large traveltime because of the power tunnel. The zero-offset time delay has been removed for (a) and (b).

3.2.2 Reconstruction test for the power tunnel model with noisy data and more accurate initial models

Here, we design a reverse test for the field detection case of the power tunnel below, in other words, this can be regarded as a credibility verification for the inverted results of field data. To be specific, the synthetic test and the field case share the same observation system setting (Fig. 14), and the relative parameter models in the synthetic test are designed according to the inverted results of field data. There are 24 *Z*-component receivers deployed on the free-surface with the spacing at 1 m, the number of sources is 13 (Fig. 14), exciting on the free-surface and loading on the *Z*-component, the source wavelets of all shots are set as the 20 Hz Ricker wavelet with the delaying time of 50 ms, thus generating 13 shot gathers that are dominated by Rayleigh waves. The recording-time is set as 0.6 s with the sampling rate of 8000 Hz, and the size of spatial calculation domain is 24.75 m × 14.75 m with the spatial-grid length at 0.25 m (Figs 15a, b and c). The background model of the power tunnel model is incremental, and the power tunnel is described by a blue rectangular abnormal body (Figs 15a, b and c). The ratio between the *P*-wave velocity and *S*-wave velocity of background model is set as 2.0, considering the lithologic difference, the ratio value V_P/V_S for the rectangular abnormal body is taken as 3.0.

The initial guess for the reconstruction test is set as the background model of the power tunnel model, the inverted *S*-wave velocity, *P*-wave velocity and density are shown by Figs 15(d), (e) and (f). The shape and position of the critical power tunnel can be delineated from the inversion images of *S*-wave velocity and density, while it is hardly reflected by the inverted result of *P*-wave velocity, clearly (Figs 15d, e and f). Similarly, the inversion system has the strongest sensitivities to *S*-wave velocity and density, the poorest sensitivity to *P*-wave velocity can be derived, again (Fig. 15). For the description of the power tunnel, the inverted *S*-wave velocity reflects an insufficient parameter estimation with a certain degree (Fig. 15d), while the inverted density expresses an overestimation (Fig. 15f). It can also be found that the model updating caused by inversion not only updates the parameters at the position of power tunnel, but also perturbs the parameters of other areas (Figs 15d, e and f), which is explained as the inherent diffusion properties of the inversion system. The abnormal body can also be reflected by the inversion image of V_P/V_S, furthermore, the characteristic related to the abnormal body is more obvious compared with the inversion images of *S*-wave velocity, *P*-wave velocity and density (Figs 15d, e and f; Fig. 16). This shows that the ratios between the *P*-wave velocity and *S*-wave velocity have been significantly updated through inversion, and for field data case, the V_P/V_S image can also be regarded as an important indicator for the anomaly inference and recognition. The difference between the waveforms calculated from the background model and observed waveforms is very small (Fig. 17a), but there is still a considerable rectangular body difference in the model domain, which prompts the necessity of using FWI/WI. The single waveform fit at the offset of −10.5 m of 7th shot shows that the scattered wave in the main event contains the core information related to the critical abnormality of underground medium (Fig. 17b).

4 FIELD DATA CASE

4.1 Analysis workflow for the field data and data characteristic

According to the above description, we give the following analysis workflow for the field data, as shown by Fig. 18. The multiscale strategy that can effectively avoid the inversion to fall into local minima (Bunks *et al*. 1995), which is also adopted for our algorithm framework, and the PSD plays a key role for optimizing the misfit function of our presented 2-D MSWI (Fig. 18).

The field seismic-wave data sets are collected above a section of the power tunnel in Changsha city, Hunan province, China (Fig. 19a) and the goal is to detect the buried position of the tunnel. According to the available information, the width and subface depth of the tunnel are about 4.1 and 7 m, respectively, the tunnel has been buried for more than 2 yr by now, and the shallow-surface medium mainly includes the loose deposits, silty clay and strong weathering slate. Therefore, we design the observed system as shown by Fig. 14 for the survey, and Figs 19(a)–(c) exhibit the survey area in detail. For the data acquiring, the 4.5 Hz *Z*-component receivers are chosen to record the seismic waves excited by a 10-kg sledgehammer, vertically (Fig. 19b). Figs 20(a) and (b) give the normalized seismograms for the sources excited at left and right, respectively, the Rayleigh-wave events still occupy most of the energy of the records,

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

the relatively strong scattered waves are also formed, and the amplitude of the refracted waves is the smallest. It can be noticed that the appearance distance of the refracted waves in Figs 20 (a) and (b) is obviously different, which indicates the lateral heterogeneity of the underground media. Figs 20(c) and (d) give the normalized amplitude spectra of the frequency-domain for the records of Figs 20(a) and (b), respectively, the peak frequency of the records is about 30–40 Hz and the dominant frequency-band is about 5–50 Hz. The spectral energy below about 10 Hz in Figs 20(c) and (d) is relatively weak, which is most likely due to the use of a 10-kg sledgehammer source in the field data experiment even though 4.5 Hz receivers are used during data acquisition. Only the fundamental-mode Rayleigh-wave is significantly reflected in the dispersion spectra generated by the phase-shift method, and the phase velocity gradually changes from 230 $\mathrm{m\,s^{-1}}$ at 10 Hz to 160 $\mathrm{m\,s^{-1}}$ at 60 Hz (Figs 20e and f). It can be found that there are obvious phase velocity deviations appear at the high-frequency band of the corresponding dispersion spectrum (Figs 20e and f), which may be also an indicator about the lateral heterogeneity of the underground media.

4.2 Inversion and geological verification

Different from the synthetic waveforms, the measured seismic waveforms are excited by a point source in the 3-D space, and an implicit condition for the degeneration from the 3-D wave equation to the 2-D wave equation is the line source configuration. Therefore, before proceeding the inversion we presented, the field data needs to be transformed from the 3-D to 2-D space to match the 2-D wave equation of *P*-SV type of elastic media (Groos *et al.* 2014), including each trace of the data is convolved with $t^{-0.5}$ (t represents the time) to achieve the phase correction, the waveforms with the offset greater than 8 m are multiplied by $r(2/t)^{0.5}$ (r denotes the offset from source to receiver) to attain the amplitude compensation. Then, a delay of 0.05 s is operated on all waveforms to prevent the non-causal parts of the source estimations generated by the source correction filter. Finally, we filter the part of the data with frequency less than 5 Hz to suppress the response from the deep stratum.

As described above, we can construct the initial *S*-wave velocity model for the 2-D MSWI based on the JSD through analysing the dispersion spectrum of the field data (Fig. 20e), as shown by Fig. 21(a). Specifically, the initial model for *S*-wave velocity is set as the 1-D incremental structure with a constant gradient value, changing from 150 $\mathrm{m\,s^{-1}}$ at the depth of 0 m to 297.5 $\mathrm{m\,s^{-1}}$ at the depth of 14.75 m (Fig. 21a), and the initial *P*-wave velocity has the same structure as the initial *S*-wave velocity with the ratio of V_P/V_S at 2.0 (Fig. 21b). The initial density is also set as an incremental structure according to the existing prior information, which changes from 1200 $\mathrm{kg\,m^{-3}}$ at the depth of 0 m to 1790 $\mathrm{kg\,m^{-3}}$ at the depth of 14.75 m (Fig. 21c). The multiscale strategy is also adopted to reduce the possibility of falling into the local minima, and we divide the entire inversion into two scales, which are 5–20 Hz and 5–50 Hz. The red rectangles in Figs 21(d)–(i) depict the position of the power tunnel given by the excavation result. For the inverted results of the first scale, there is an obvious low-density perturbation appearing in the left part of the inversion image of density (Fig. 21f), which is an indicator about the power tunnel, while the tunnel is not directly reflected from the inversion images of *S*-wave velocity and *P*-wave velocity (Figs 21d and e). It can be found that the size of the low-density area in the left area is greater than that of the tunnel (Figs 21f and i), which may show that the medium near the tunnel is still not completely compacted after 2 yr. The inverted *S*-wave velocity of the second scale have a clearer expression for the power tunnel compared with that of the first scale (Fig. 21g). The *P*-wave velocity belonging to the tunnel position of the second scale is higher than that of the first scale (Figs 21e and h), which reveals the high V_P/V_S values about the tunnel position.

In addition, we also show the inversion image of the ratios between *P*-wave velocity and *S*-wave velocity (Fig. 22). The V_P/V_S image implements a more direct description for the position of the power tunnel compared with other inversion images (Figs 21d, e, f, g, h, i and 22). The significant higher ratios between *P*-wave velocity and *S*-wave velocity occur at the tunnel position (Fig. 22), which is a direct response for the internal caves of the tunnel. Although the burial position of the tunnel has been verified by related images (Figs 21g, h, i and 22), we still collect the shallow-surface geological information near the survey area to inspect the reasonability of the inverted results. Table 1 shows the detailed description about the borehole data near the survey area, and the underground medium is divided into three formations, consisting of the loose deposits, silty clay and strong weathering slate, respectively. It can be noted that the interface between the first and second layer is laterally variable, with a depth range of 3.0–7.0 m (Table 1), as the inverted *S*-wave velocity and *P*-wave velocity reflected (Figs 21g and h). The approximate *S*-wave velocity values of the corresponding formation are also derived from the rock handbook, on the whole, the value range of the inverted *S*-wave velocity is consistent with that of the borehole data (Fig. 21g and Table 1), which provides a proof for the reliability. It can be noted that the inverted *S*-wave velocity and *P*-wave velocity have a relatively consistent structure (Figs 21g and h), while the structure of the inverted density is significantly different from that of these two parameters (Figs 21g, h and i). These significant differences are unlikely to be illusions caused by the crosstalk of parameter classes, but reflecting the original differences between different parameter classes, because the approximate term of the inverse Hessian operator in the framework of our method can suppress the crosstalk to a certain extent.

Also, the inverted results are evaluated from the perspective of data domain (Figs 23, 24 and 25). The source estimations at the 0th iteration of the first scale have similar phases, ensuring the data coherence between different shot gather (Fig. 23a). The source estimations at the final iteration of the first scale have been updated to a certain extent due to the strategy of dynamic update settings at each iteration, especially the estimated source wavelet of the 6th shot (Fig. 23a). After the multiscale inversion, the misfit function deserves a considerable reduction, decreasing to 0.3 of the initial misfit value (Fig. 23b). Figs 24(a), (c) and (e) show the comparison between the observed and inverted waveforms at the second scale for the first, 7th and 13th shot, as a whole, the observed waveforms have been properly matched by the inverted waveforms with small discrepancies (Figs 24a, b, c, d, e and f). For the near-source channels of 7th and 13th shot, the residuals between the observed and inverted waveforms are relatively large (Figs 24d and f), which may reflect the artificial noise during the source excitement and possible near-source effects. The near-source channels account for a small proportion of the entire observed data, therefore a considerable adverse effect to the inversion would be not generated. Fig. 25 show the common-offset profiles of the observed data and predicted data of the second scale, and these two profiles are similar, especially the main event. The significant large traveltime features in the left part of the observed common-offset profile reflect the relatively low average *S*-wave velocity caused by the power tunnel, as indicated

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

by the black arrow (Fig. 25a). Even, we can determine the existence of the power tunnel and identify its horizontal position according to the large traveltime features (Fig. 25a). Also, the traveltime curve of the observed common-offset profile can be accurately matched by the main event of the predicted common-offset profile (Fig. 25b), which illustrates the reliability of the inverted results provided by our method.

5 CONCLUSION

In this paper, we present a novel 2-D MSWI based on the JSD. The novel approach achieves an implicit weighting along the time-axis for each 1-D adjoint source of the CWI, thus enhancing the extra illumination for deeper medium. The synthetic tests show that under the same conditions, the reconstruction accuracy of the *S*-wave velocity of our presented method is increased by about 30 per cent compared with the CWI, and the density reconstruction performance is sometimes improved by about 20 per cent. Furthermore, our method manifests the resilience to white Gaussian noise in the data. We also test the novel method by a field data case, and the buried position of the tunnel is successfully verified from the generated tomograms. Regardless of the synthetic tests or field data case, the inversion system has almost the strongest sensitivities to the *S*-wave velocity and density, performing the poorest sensitivity to the *P*-wave velocity. Last but not least, the same method for the 2-D wave equation of SH type using Love waves in observed data can also be implemented along our path.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (Grant No. U1901602), Shenzhen Science and Technology Program (Grant No. KQTD20170810111725321), Guangdong Provincial Key Laboratory of Geophysical High-resolution Imaging Technology (2022B1212010002) and Shenzhen Key Laboratory of Deep Offshore Oil and Gas Exploration Technology (Grant No. ZDSYS20190902093007855). This work is also supported by the National Natural Science Foundation of China (Grant No. 42222407). The author contribution of this research is stated as follows.

AUTHOR CONTRIBUTIONS

Yingwei Yan: conceptualization, methodology, software, investigation, writing—original draft, writing—review and editing, visualization. Xiaofei Chen: Formal analysis, funding acquisition. Jing Li: validation, project administration. Jianbo Guan: supervision. Yu Li: data curation. Shihao Cui: supervision.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

Data and codes can be available by contacting the corresponding author.

REFERENCES

Aster, R., Borchers, B. & Thurber, C., 2018. *Parameter Estimation and Inverse Problems,* Vol. **III,** Elsevier.

Bertsekas, D.P., 1982. *Constrained Optimization And Lagrange Multiplier Methods,* Academic Press.

Borisov, D., Modrak, R., Gao, F. & Tromp, J., 2018. 3D elastic full-waveform inversion of surface waves in the presence of irregular topography using an envelope-based misfit function. *Geophysics,* **83**(1), R1–R11.

Bunks, C., Saleck, F., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion. *Geophysics,* **60**(5), 1457–1473.

Chen, J., Pan, L., Li, Z. & Chen, X., 2022. Continental reworking in the eastern South China Block and its adjacent areas revealed by F-J multimodal ambient noise tomography. *J. geophys. Res.,* **127,** e2022JB024776.

Choi, Y. & Alkhalifah, T., 2012. Application of multi-source waveform inversion to marine streamer data using the global correlation norm. *Geophys. Prospect.,* **60**(4), 748–758.

Crase, E., Pica, A., Noble, M., McDonald, J. & Tarantola, A., 1990. Robust elastic nonlinear waveform inversion: application to real data. *Geophysics,* **55**(5), 527–538.

Dokter, E., Köhn, D., Wilken, D., De Nil, D. & Rabbel, W., 2017. Full-waveform inversion of SH- and Love-wave data in near-surface prospecting. *Geophys. Prospect.,* **65 (** 1), 216–236.

Engquist, B. & Yang, Y., 2022. Optimal transport based seismic inversion: beyond cycle skipping. *Commun. Pure Appl. Math.,* **75**(10), 2201–2244.

Fan, Y., Chen, X., Liu, X., Liu, J. & Chen, X., 2007. Approximate decomposition of the dispersion equation at high frequencies and the number of multimodes for Rayleigh waves. *Chinese J. Geophys. (in Chinese),* **50**(1), 222–230.

Farrugia, D., Paolucci, E., D'Amico, S. & Galea, P., 2016. Inversion of surface wave data for subsurface shear wave velocity profiles characterized by a thick buried low-velocity layer. *Geophys. J. Int.,* **206**(2), 1221–1231.

Feng, S., Sugiyama, T. & Yamanaka, H., 2005. Effectiveness of multimode surface wave inversion in shallow engineering site investigations. *Explor. Geophys.,* **36**(1), 26–33.

Foti, S. *et al.* 2018. Guidelines for the good practice of surface wave analysis: a product of the InterPACIFIC project. *Bull. Earthq. Eng.,* **16,** 2367–2420.

Foti, S., Lai, C., Rix, G. & Strobbia, C., 2014. *Surface Wave Methods for Near-Surface Site Characterization,* CRC Press.

Gao, L., Pan, Y. & Bohlen, T., 2020. 2-D multiparameter viscoelastic shallow-seismic full-waveform inversion: reconstruction tests and first field-data application. *Geophys. J. Int.,* **222**(1), 560–571.

Gao, L., Pan, Y., Rieder, A., Bohlen, T. & Mao, W., 2023. Multiparameter 2-D viscoelastic full-waveform inversion of Rayleigh waves: a field experiment at Krauthausen test site. *Geophys. J. Int.,* **234**(1), 297–312.

Groos, L., Schäfer, M., Forbriger, T. & Bohlen, T., 2014. The role of attenuation in 2D full-waveform inversion of shallow-seismic body and Rayleigh waves. *Geophysics,* **79**(6), R247–R261.

Groos, L., Schäfer, M., Forbriger, T. & Bohlen, T., 2017. Application of a complete workflow for 2D elastic full-waveform inversion to recorded shallow-seismic Rayleigh waves. *Geophysics,* **82**(2), R109–R117.

Guan, J., Li, Y., Ji, R., Liu, G. & Yan, Y., 2022. Love wave full-waveform inversion for archaeogeophysics: from synthesis tests to a field case. *J. appl. Geophys.,* **202,** doi:10.1016/j.jappgeo.2022.104653.

Hu, S., Zhao, Y., Socco, L.V. & Ge, S., 2021. Retrieving 2-D laterally varying structures from multistation surface wave dispersion curves using multiscale window analysis. *Geophys. J. Int.,* **227**(2), 1418–1438.

Köhn, D. *et al.* 2018. Seismic SH full waveform inversion as new prospection method in archaeogeophysics, in *Proceedings of the 24th European Meeting of Environmental and Engineering Geophysics,* Porto, Portugal. .

Lamuraglia, S., Stucchi, E. & Aleardi, M., 2023. Application of a global–local full-waveform inversion of Rayleigh wave to estimate the near-surface shear wave velocity model. *Near Surf. Geophys.,* **21**(1), 21–38.

Li, J., Feng, Z. & Schuster, G., 2017. Wave-equation dispersion inversion. *Geophys. J. Int.,* **208**(3), 1567–1578.

Lin, J., 1991. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory,* **37**(1), 145–151.

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

Liu, J., Ghose, R. & Draganov, D., 2022. Characterizing near-surface structures at the Ostia archaeological site based on instantaneous-phase coherency inversion. *Geophysics,* **87**(4), R337–R348.

Liu, Y., Zhang, Y., Nilot, E., Yannick, C.H.N. & Yang, P., 2021. Detection of a shallow-buried rock obstruction using 2D full waveform inversion. *Soil Dyn. Earthq. Eng.,* **143,** doi:10.1016/j.soildyn.2021.106644.

Luo, Y., Xia, J., Miller, R.D., Xu, Y., Liu, J. & Liu, Q., 2009. Rayleigh-wave mode separation by high-resolution linear Radon transform. *Geophys. J. Int.,* **179**(1), 254–264.

Manning, C. & Schütze, H., 1999. *Foundations of Statistical Natural Language Processing,* MIT Press.

Maraschini, M. & Foti, S., 2010. A Monte Carlo multimodal inversion of surface waves. *Geophys. J. Int.,* **182**(3), 1557–1566.

Mario, B. & Patrizia, B., 1998. *Introduction to Inverse Problems in Imaging,* Vol. **I,** CRC Press.

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

Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multi-offset seismic data. *Geophysics,* **52**(9), 1211–1228.

Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization,* Springer.

Operto, S., Virieux, J., Dessa, J.X. & Pascal, G., 2006. Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern Nankai trough. *J. geophys. Res.,* **111**(B9).

Pan, Y., Gao, L. & Bohlen, T., 2019. High-resolution characterization of near-surface structures by surface wave inversions: from dispersion curve to full waveform. *Surv. Geophys.,* **40,** 167–195.

Pan, Y., Xia, J., Xu, Y., Gao, L. & Xu, Z., 2016. Love-wave waveform inversion for shallow shear-wave velocity. *Geophysics,* **81**(1), R1–R14.

Park, C.B., Miller, R.D. & Xia, J., 1998. Imaging dispersion curves of surface waves on multi-channel record, in *Proceedings of the SEG Technical Program Expanded Abstracts,* pp. 1377–1380.

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**(2), 495–503.

Pratt, R.G., Shin, C. & Hicks, G.J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophys. J. Int.,* **133**(2), 341–362.

Romdhane, A., Grandjean, G., Brossier, R., Rejiba, F. & Virieux, J., 2011. Shallow-structure characterization by 2D elastic full-waveform inversion. *Geophysics,* **76**(3), R81–R93.

Shannon, C.E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.,* **27**(4), 623–656.

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.

Snieder, R. & Trampert, J., 1999. Inverse problems in geophysics, in *Wavefield Inversion. International Centre for Mechanical Sciences,* Vol. **398,** ed. Wirgin, A., Springer.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation. *Geophysics,* **49**(8), 1259–1266.

Tavares, D.M. & Lucena, L.S., 2005. Entropy analysis of stochastic processes at finite resolution. *Physica A,* **357**(1), 71–78.

Tran, K.T., Mcvay, M., Faraone, M. & Horhota, D., 2013. Sinkhole detection using 2D full seismic waveform tomography. *Geophysics,* **78**(5), R175–R183.

Vantassel, J.P. & Cox, B.R., 2021. SWinvert: a workflow for performing rigorous 1-D surface wave inversions. *Geophys. J. Int.,* **224**(2), 1141–1156.

Vigh, D., Jiao, K., Watts, D. & Sun, D., 2014. Elastic full-waveform inversion application using multicomponent measurements of seismic data collection. *Geophysics,* **79**(2), R63–R77.

Wang, Y., 2017. *Seismic Inversion: Theory And Applications,* John Wiley & Sons.

Wang, Z., Sun, C. & Wu, D., 2023. Near-surface site characterization based on joint iterative analysis of first-arrival and surface-wave data. *Surv. Geophys.,* **44,** 357–386.

Wu, R., Luo, J. & Wu, B., 2014. Seismic envelope inversion and modulation signal model. *Geophysics,* **79**(3), WA13–WA24.

Xia, J., Gao, L., Pan, Y., Shen, C. & Yin, X., 2015. New findings in high-frequency surface wave method. *Chinese J. Geophys. (in Chinese),* **58**(8), 2591–2605.

Yan, Y., Chen, X., Huai, N. & Guan, J., 2022b. Modern inversion workflow of the multimodal surface wave dispersion curves: staging strategy and Pattern search with embedded Kuhn–Munkres algorithm. *Geophys. J. Int.,* **231**(1), 47–71.

Yan, Y., Li, J., Huai, N., Guan, J. & Liu, H., 2022a. Two-station analysis of passive surface waves with continuous wavelet transform and plane-wave-based beamforming. *J. appl. Geophys.,* **197,** doi:10.1016/j.jappgeo.2021.104526.

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

APPENDIX A: DERIVATION OF GRADIENT AND QUASI-HESSIAN OPERATOR

Before the derivation, we first give the convention of vector-to-vector differentiation, which is as follows:

$$\begin{cases} \frac{\partial(\mathbf{Gx})}{\partial \mathbf{x}} = \mathbf{G}^T, \\ \frac{\partial(\mathbf{x}^\mathbf{T}\mathbf{G})}{\partial \mathbf{x}} = \mathbf{G}. \end{cases} \tag{A1}$$

Where $\mathbf{x}$ is a vector, the matrix $\mathbf{G}$ is not a function of $\mathbf{x}$, and the symbol T represents the transpose operation.

Generally, the wave equations to be solved are non-homogeneous, consisting of the specific partial-differential operator, initial and boundary conditions. Particularly, we consider the first-order wave equation of P-SV type for 2-D elastic media, whose abstract form is:

$$\begin{cases} S(p)\,u(\mathbf{X}, t) = \Phi, \\ u(\mathbf{X}, 0) = 0, \\ u(\infty, t) = 0, \\ u(\mathbf{X}, t)|_{z=0} = f_r. \end{cases} \tag{A2}$$

In the above formula, p represents the physical parameter, which are parametrized as the S-wave velocity, P-wave velocity and density in the paper. $S(p)$ denotes the partial-differential operator related to the parameter p, also called as the forward operator of 2-D wave equation of P-SV type in elastic media. $\mathbf{X}$ is a vector that describes the space position, t represents the time (a finite number), u denotes the wavefield solved from the wave equation, Φ is the source term and f_r represents the free-surface boundary condition. To simplify the derivation, the space-position vector $\mathbf{X}$ and boundary conditions in the wave equation would be omitted below. The specific form of the partial-differential operator $S(p)$ is as follows:

$$\begin{aligned} S(p) &= \begin{bmatrix} \rho\partial_t & 0 & -\partial_x & 0 & -\partial_z \\ 0 & \rho\partial_t & 0 & -\partial_z & -\partial_x \\ -(\lambda+2\mu)\,\partial_x & -\lambda\partial_z & \partial_t & 0 & 0 \\ -\lambda\partial_x & -(\lambda+2\mu)\,\partial_z & 0 & \partial_t & 0 \\ -\mu\partial_z & -\mu\partial_x & 0 & 0 & \partial_t \end{bmatrix} \\ &= \begin{bmatrix} \rho & 0 & 0 & 0 & 0 \\ 0 & \rho & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \partial_t + \begin{bmatrix} 0 & 0 & -\partial_x & 0 & -\partial_z \\ 0 & 0 & 0 & -\partial_z & -\partial_x \\ -(\lambda+2\mu)\,\partial_x & -\lambda\partial_z & 0 & 0 & 0 \\ -\lambda\partial_x & -(\lambda+2\mu)\,\partial_z & 0 & 0 & 0 \\ -\mu\partial_z & -\mu\partial_x & 0 & 0 & 0 \end{bmatrix} \\ &= A\partial_t + B, \end{aligned} \tag{A3}$$

where A is a coefficient matrix related to the partial derivative for t, and B represents the part related to the spatial partial derivative. ρ is the density, λ has the form of $\lambda = \rho V_P{}^{2-}2\rho V_S{}^2$, and μ is the shear modulus with the form of $\mu = \rho V_S{}^2$. The detailed form of the wavefield u is:

$$u = (v_x, v_z, \sigma_{xx}, \sigma_{zz}, \sigma_{xz})^T. \tag{A4}$$

Where v_x and v_z represent the particle velocity in the time domain for the x-direction and z-direction, respectively, σ_{xx}, σ_{zz} and σ_{xz} refer to the stress in the time domain.

For the FWI or WI, the most frequently used misfit function is derived from the L^2 norm. Meanwhile, we propose a novel misfit function based on the Jensen–Shannon divergence (JSD) in the paper. Here, only the inversion system with the single shot is considered, for multiple shot settings, the misfit function, gradient and quasi-Hessian operator can be computed through stacking them for each shot, simply. Once the misfit function is constructed, the significant thing is to calculate the first-order derivatives (gradient vector) and approximate second-order derivatives (quasi-Hessian matrix) of the misfit function to proceed the local optimization. Note that the wave field u is an implicit function of the parameter p and the source term Φ, therefore the misfit function E is also an implicit function about p and Φ, which is formulated as:

$$E(p, \Phi) = f(u(p, \Phi)) = f(u, p, \Phi). \tag{A5}$$

It is clearly stated here that p and Φ are independent in the above formula. Since we use the source-correction filter to update the source term Φ during the inversion process, only the derivation for the gradient and quasi-Hessian operator of the misfit function with respect to the parameter p is given in the following part, and the gradient is expressed as:

$$\frac{\partial E(p, \Phi)}{\partial p} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial p} = \left(\frac{\partial u}{\partial p}\right)^T \frac{\partial f}{\partial u} = J^T \frac{\partial f}{\partial u}, \tag{A6}$$

where J is the Jacobian matrix, also named as the partial-differential wavefield in some scenarios. Because of the huge scale of the problem, it is difficult to calculate the gradient by finite difference, but we can regard it as a constraint optimization problem, which can be solved by the Lagrange multiplier method, and the specific form of the Lagrange functionality F is as follows:

$$\begin{aligned} F\left(u, p, \Phi, \lambda^*, \mu^0\right) &= f(u, p, \Phi) + \int_0^{t_C} \langle \lambda^*(t), \varphi(u, p)\rangle_\Omega \mathrm{d}t + \left\langle \mu^0, u(0)\right\rangle_\Omega \\ &= f(u, p, \Phi) + \int_0^{t_C} \langle \lambda^*(t), S(p)\,u(t) - \Phi(t)\rangle_\Omega \mathrm{d}t + \left\langle \mu^0, u(0)\right\rangle_\Omega. \end{aligned} \tag{A7}$$

Among them, λ^* and μ^0 are the introduce parameter, t_C is the total recording time, and $u(0)$ refers to the wavefield at $t = 0$. Ω represents the space integral domain, whose specific meaning is:

$$\langle \lambda^*(t), u(t)\rangle_\Omega = \lambda^*(\Omega, t)^T u(\Omega, t). \tag{A8}$$

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

$\lambda^*(\Omega, t)$ and $u(\Omega, t)$ are the infinite-dimensional vector at the time t covered with the entire space integral domain Ω. The form of the constraint condition $\varphi(u, p)$ in the formula (A7) is:

$$\varphi(u, p) = S(p)u(t) - \Phi(t) = (A\partial_t + B)u(t) - \Phi(t) = 0. \tag{A9}$$

Utilizing the integral subsection method, we can get:

$$\begin{aligned}
&\int_0^{t_C} \langle \lambda^*(t), S(p)u(t) - \Phi(t)\rangle_\Omega \mathrm{d}t = \int_0^{t_C} \langle \lambda^*(t), A\partial_t u(t)\rangle_\Omega \mathrm{d}t + \int_0^{t_C} \langle \lambda^*(t), Bu(t)\rangle_\Omega \mathrm{d}t \\
&- \int_0^{t_C} \langle \lambda^*(t), \Phi(t)\rangle_\Omega \mathrm{d}t \\
&= \langle \lambda^*(t_C), Au(t_C)\rangle_\Omega - \langle \lambda^*(0), Au(0)\rangle_\Omega - \int_0^{t_C} \langle \partial_t\lambda^*(t), Au(t)\rangle_\Omega \mathrm{d}t + \int_0^{t_C} \langle \lambda^*(t), Bu(t)\rangle_\Omega \mathrm{d}t \\
&- \int_0^{t_C} \langle \lambda^*(t), \Phi(t)\rangle_\Omega \mathrm{d}t \\
&= \lambda^*(\Omega, t_C)^T Au(\Omega, t_C) - \lambda^*(\Omega, 0)^T Au(\Omega, 0) - \int_0^{t_C} (\partial_t\lambda^*(\Omega, t))^T Au(\Omega, t)\,\mathrm{d}t + \\
&\int_0^{t_C} \lambda^*(\Omega, t)^T Bu(\Omega, t)\mathrm{d}t - \int_0^{t_C} \lambda^*(\Omega, t)^T \Phi(\Omega, t)\mathrm{d}t
\end{aligned} \tag{A10}$$

Let the first-order partial derivative of formula (A7) with respect to u equal to 0, which is expressed as follows:

$$\frac{\partial F}{\partial u}(t) = \begin{cases} -A^T\lambda^*(t) + \frac{\partial f}{\partial u}(t) - \left(A^T\partial_t - B^T\right)\lambda^*(t) + \mu^0 = 0 \; t = 0, \\ \frac{\partial f}{\partial u}(t) - \left(A^T\partial_t - B^T\right)\lambda^*(t) = 0 \; t \in (0, t_C), \\ A^T\lambda^*(t) + \frac{\partial f}{\partial u}(t) - \left(A^T\partial_t - B^T\right)\lambda^*(t) = 0 \; t = t_C. \end{cases} \tag{A11}$$

Please note that the symbol Ω for the spatial integration domain in the above formula has been omitted. Although the above expression is given by the segmented function form, it is still a continuous function. After the simplification, we can get the following expression:

$$\begin{cases} \mu^0 = A^T\lambda^*(t), t = 0, \\ \left(A^T\partial_t - B^T\right)\lambda^*(t) = \frac{\partial f}{\partial u}(t), \; t \in [0, t_C], \\ \lambda^*(t) = 0, \; t = t_C. \end{cases} \tag{A12}$$

Because the introduce parameter λ^* equals to 0 at the time t_C, the right-hand term in the second row needs to be reverse-time propagated to compute the λ^*. For certain contexts, the introduce parameter λ^* is also called as the adjoint wavefield, and the corresponding wave equation is also named as the adjoint wave equation. Please note the adjoint wave equation is not the transposition of the forward wave equation, and the additional negative sign is added to the spatial partial derivatives. The right-hand term in the second row describes the relationship between the adjoint source and misfit function, meaning that the different adjoint sources and wavefields would be generated due to the different definitions of misfit functions. Similarly, for the formula (A7), let the first-order partial derivative about the parameter p equal to 0, which is determined as:

$$\frac{\partial F}{\partial p} = J^T\frac{\partial f}{\partial u} + \int_0^{t_C} \left\langle \frac{\partial S(p)}{\partial p}u(t), \lambda^*(t)\right\rangle \mathrm{d}t = \mathbf{0} \; t \in [0, t_C]. \tag{A13}$$

Then, the gradient of the misfit function with respect to a specific parameter p_i can be derived by the equivalent constraint above, which is expressed as:

$$\frac{\partial E}{\partial p}\Big|_{p=p_i} = -t_c\left\langle \frac{\partial S(p)}{\partial p}u, \lambda^*\right\rangle\Big|_{p=p_i} \; i = 1, 2, \ldots, M. \tag{A14}$$

Where M represents the number of the parameters of the inversion system. According to the above formula, only an extra forward simulation is needed for the calculation of the gradient vector of the misfit function, and this high-efficiency computation method is called as the adjoint-state method in some contexts. t_C is a constant, and only the direction cosine of the gradient vector is needed for our method, therefore we can simplify the gradient vector of the misfit function through the following expression:

$$\frac{\partial E}{\partial p}\Big|_{p=p_i} = -\left\langle \frac{\partial S(p)}{\partial p}u, \lambda^*\right\rangle\Big|_{p=p_i} \; i = 1, 2, \ldots, M. \tag{A15}$$

It can be seen from the formula (A15) that the calculation of the gradient vector of the misfit function is related to the adjoint wavefield, which means that different gradient vectors would be formed for different misfit function settings. Let the first-order partial derivative of formula (A7) with respect to the introduce parameter λ^* and μ^0 equal to 0, we can get the following relational expression:

$$\begin{cases} \frac{\partial F}{\partial \lambda^*}(t) = S(p)u(t) - \Phi(t) = 0 \; t \in [0, t_C], \\ \frac{\partial F}{\partial \mu^0} = u(0) = 0. \end{cases} \tag{A16}$$

It can be found that, let the first-order partial derivative of formula (A7) with respect to u equal to 0, we can get the adjoint wave equation that constrains the adjoint wavefield λ^*; let the first-order partial derivative of formula (A7) with respect to the adjoint wavefield λ^* and μ^0 equal to 0, we can get the wave equation that constrains the wavefield u. Because of this property, the wavefield u and adjoint wavefield λ^* are called as the dual quantity.

Here, we use the approximate $J^{\mathrm{T}}J$ as an estimation of the Hessian matrix of the misfit function. According to the formula (A12) to expand the formula (A13), we can obtain the expression as follows:

$$\frac{\partial F}{\partial p} = \left\langle J, \left(A^T\partial_t - B^T\right)\lambda^*(t)\right\rangle + \int_0^{t_C} \left\langle \frac{\partial S(p)}{\partial p}u(t), \lambda^*(t)\right\rangle \mathrm{d}t = \mathbf{0} \; t \in [0, t_C]. \tag{A17}$$

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

Ignoring the public item $\lambda^*(t)$ in the above formula, the following expression is obtained:

$$\begin{aligned}&\left\langle \frac{\partial S(p)}{\partial p}u,\ \frac{\partial S(p)}{\partial p}u\right\rangle = \left\langle -\frac{\partial S(p)}{\partial p}u,\ -\frac{\partial S(p)}{\partial p}u\right\rangle \\ &= \frac{1}{t_c{}^2}J^T\left(A^T\partial_t - B^T\right)(A\partial_t - B)\,J.\end{aligned} \tag{A18}$$

Formula (A18) realizes a filter to the $J^{\mathrm{T}}J$, which can be regarded as the approximate $J^{\mathrm{T}}J$, therefore the quasi-Hessian operator H of the misfit function can be expressed as follows:

$$H_{ij} = \left\langle \frac{\partial S(p)}{\partial p}u(p)\big|_{p=p_i},\ \frac{\partial S(p)}{\partial p}u(p)\big|_{p=p_j}\right\rangle\ i = 1, 2, \ldots, M;\ j = 1, 2, \ldots, M. \tag{A19}$$

It can be noted that the calculation of quasi-Hessian operator is independent of the adjoint wavefield. In other words, the quasi-Hessian operator in the CWI and our method is exactly the same in the paper, which means that the performance improvement of our method benefits from the misfit function derived from the JSD. The matrix H is a diagonally dominant matrix under the high-frequency approximation (Pratt *et al.* 1998); therefore, we only calculate the diagonal elements of the quasi-Hessian operator through the above formula. For different first-order partial-differential operators and misfit functions, the mathematical form of the gradient, quasi-Hessian operator and adjoint source remains invariable. When the first-order partial-differential operator $S(p)$ is the first-order wave equation of P-SV type for 2-D elastic media, the detailed expressions of a certain element for the gradient vector and quasi-Hessian operator (only consisting of the diagonal elements) are as follows:

$$\begin{cases}\frac{\partial E}{\partial V_S} = -2\rho V_S\sum\limits_{n=1}^{nt}\left(2\frac{\partial v_z{}^{(n)}}{\partial z}\sigma_{xx}{}^{*(n)} + 2\frac{\partial v_x{}^{(n)}}{\partial x}\sigma_{zz}{}^{*(n)} - \frac{\partial v_x{}^{(n)}}{\partial z}\sigma_{xz}{}^{*(n)} - \frac{\partial v_z{}^{(n)}}{\partial x}\sigma_{xz}{}^{*(n)}\right),\\ \frac{\partial E}{\partial V_P} = 2\rho V_P\sum\limits_{n=1}^{nt}\left(\frac{\partial v_x{}^{(n)}}{\partial x} + \frac{\partial v_z{}^{(n)}}{\partial z}\right)\left(\sigma_{xx}{}^{*(n)} + \sigma_{zz}{}^{*(n)}\right),\\ \frac{\partial E}{\partial \rho} = -\sum\limits_{n=1}^{nt}\left(\frac{\partial v_x{}^{(n)}}{\partial t}v_x{}^{*(n)} + \frac{\partial v_z{}^{(n)}}{\partial t}v_z{}^{*(n)}\right) + 2V_S{}^2\left(\frac{\partial v_z{}^{(n)}}{\partial z}\sigma_{xx}{}^{*(n)} + \frac{\partial v_x{}^{(n)}}{\partial x}\sigma_{zz}{}^{*(n)}\right) -\\ V_P{}^2\left(\frac{\partial v_x{}^{(n)}}{\partial x} + \frac{\partial v_z{}^{(n)}}{\partial z}\right)\left(\sigma_{xx}{}^{*(n)} + \sigma_{zz}{}^{*(n)}\right) - V_S{}^2\left(\frac{\partial v_x{}^{(n)}}{\partial z} + \frac{\partial v_z{}^{(n)}}{\partial x}\right)\sigma_{xz}{}^{*(n)}.\end{cases} \tag{A20a}$$

$$\begin{cases}H(V_S) = \sum\limits_{n=1}^{nt}\left[16\rho^2V_S{}^2\frac{\partial v_z{}^{(n)}}{\partial z}\frac{\partial v_z{}^{(n)}}{\partial z} + 16\rho^2V_S{}^2\frac{\partial v_x{}^{(n)}}{\partial x}\frac{\partial v_x{}^{(n)}}{\partial x} + 4\rho^2V_S{}^2\left(\frac{\partial v_x{}^{(n)}}{\partial z} + \frac{\partial v_z{}^{(n)}}{\partial x}\right)\left(\frac{\partial v_x{}^{(n)}}{\partial z} + \frac{\partial v_z{}^{(n)}}{\partial x}\right)\right],\\ H(V_P) = \sum\limits_{n=1}^{nt}\left[8\rho^2V_P{}^2\left(\frac{\partial v_x{}^{(n)}}{\partial x} + \frac{\partial v_z{}^{(n)}}{\partial z}\right)\left(\frac{\partial v_x{}^{(n)}}{\partial x} + \frac{\partial v_z{}^{(n)}}{\partial z}\right)\right],\\ H(\rho) = \sum\limits_{n=1}^{nt}\frac{\partial v_x{}^{(n)}}{\partial t}\frac{\partial v_x{}^{(n)}}{\partial t} + \frac{\partial v_z{}^{(n)}}{\partial t}\frac{\partial v_z{}^{(n)}}{\partial t} + \left(V_P{}^4 + \left(V_P{}^2 - 2V_S{}^2\right)^2\right)\left(\frac{\partial v_x{}^{(n)}}{\partial x}\frac{\partial v_x{}^{(n)}}{\partial x} + \frac{\partial v_z{}^{(n)}}{\partial z}\frac{\partial v_z{}^{(n)}}{\partial z}\right) +\\ 4V_P{}^2\left(V_P{}^2 - 2V_S{}^2\right)\frac{\partial v_x{}^{(n)}}{\partial x}\frac{\partial v_z{}^{(n)}}{\partial z} + V_S{}^4\left(\frac{\partial v_x{}^{(n)}}{\partial z} + \frac{\partial v_z{}^{(n)}}{\partial x}\right)\left(\frac{\partial v_x{}^{(n)}}{\partial z} + \frac{\partial v_z{}^{(n)}}{\partial x}\right).\end{cases} \tag{A20b}$$

Where V_S, V_P and ρ are the S-wave velocity, P-wave velocity and density for a certain spatial-grid, respectively; nt is the number of the sampling points in the time domain, v_x, v_z, σ_{xx}, σ_{zz} and σ_{xz} refer to the forward wavefield of the same spatial-grid in the time domain, and the variates containing the superscript $*$ represent the adjoint wavefield of the same spatial-grid in the time domain. In short, different misfit functions would define different adjoint sources, then generating different adjoint wavefields and gradients, and the extra illumination compensation may be formed due to the different misfit functions. For the misfit function derived by the L^2 norm in the paper, the detailed expression for the adjoint source of a certain channel is as follows:

$$\Phi_{L^2}{}^* = \frac{1}{\max(|u|)}\left(u^N - d^N\right). \tag{A21}$$

Here, u is the predicted seismic signal for a certain channel, u^{N} represents the corresponding predicted signal with normalized operation and d^{N} denote the corresponding normalized seismic signal of the observed data [the normalized operation is described by the first row of formula (4)]. For the misfit function defined by the JSD, the detailed 1-D adjoint source is expressed as:

$$\Phi_{JSD}{}^* = \frac{1}{2\max(|u|)}\frac{\left\langle\frac{u}{\max(|u|)} + 1.1\right\rangle - \frac{u}{\max(|u|)} - 1.1}{\left\langle\frac{u}{\max(|u|)} + 1.1\right\rangle^2}\ln\left(\frac{2u^N}{u^N + d^N}\right). \tag{A22}$$

Here, u^{N} and d^{N} denote the predicted and observed seismic signal with normalized operation [the normalized operation is described by the second row of formula (4)], respectively; and the meaning of the symbol $<\ldots>$ is the same as the formula (5) above.

APPENDIX B: SOURCE ESTIMATION

Source estimation is an indispensable part for FWI or WI. Here, we utilize the basic principle that the seismic signal of each channel is the convolution of the source wavelet and Green's function in the time domain, and regard the source estimation as a damped least-square optimization problem, whose form is as follows:

$$\begin{cases}F(c_l;\varepsilon) = \sum\limits_{l=0}^{N-1}\sum\limits_{k=1}^{M}|d_{lk} - c_lu_{lk}|^2 + M\bar{E}\varepsilon^2|c_l|^2,\ \textit{therein}: d_{lk} = s_l{}^{Opt}G_{lk}{}^d, u_{lk} = s_lG_{lk}{}^p,\\ \bar{E} = \frac{1}{MN}\sum\limits_{l=0}^{N-1}\sum\limits_{k=1}^{M}|u_{lk}|^2,\end{cases} \tag{B1}$$

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024

where F is the objective function to be optimized, d_{lk} is the Fourier expansion coefficient at the angular frequency $\omega_l = l\Delta\omega$ of the kth channel field data in the time domain (the amplitude is normalized with respect to the maximum absolute amplitude in the time domain), u_{lk} is the corresponding Fourier expansion coefficient of the predicted data (the same normalization operation is adopted). $G_{lk}{}^{\mathrm{d}}$ and $G_{lk}{}^{\mathrm{p}}$ represent the Green's function at the angular frequency $\omega_l = l\Delta\omega$ for the kth channel of the field and predicted data, respectively; c_l is the filter coefficient, s_l is the source wavelet of the predicted data at the frequency domain, and $s_l{}^{\mathrm{Opt}}$ is the optimized source wavelet for the field data at the frequency domain that need to be estimated; ε is the damping parameter, and it is set as 10.0 for the field data case in the paper. Let the first derivative of the objective function F with respect to c_l being zero, and we can get the damped least-square solution for the filter coefficient c_l, which is as follows:

$$c_l = \frac{\sum_{k=1}^{M} d_{lk} u_{lk}{}^{*}}{M\bar{E}\varepsilon^2 + \sum_{k=1}^{M} |u_{lk}|^2}. \tag{B2}$$

Here, the superscript $*$ for u_{lk} refers to the complex conjugate operation, then we can use the coefficient c_l to filter the source wavelet s_l of the predicted data to obtain the optimized source for the field data at the frequency domain, which is expressed as:

$$s_l{}^{Opt} = c_l s_l. \tag{B3}$$

Finally, we apply the inverse Fourier transform to the optimized source wavelet in the above formula to obtain the estimated source of the field data in the time domain. Such a source estimation process is actually a filter operation to the source wavelet of the predicted data; therefore, it is called as the source correction filter (Groos *et al.* 2014).

© The Author(s) 2024. Published by Oxford University Press on behalf of The Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/gji/article/238/1/132/7651287 by Ecole Polytechnique de Montreal user on 21 May 2024