ABSTRACT

The full-waveform inversion (FWI) aims at estimating subsurface physical parameter by minimizing the misfit between simulated data and observations. FWI relies heavily on an accurate initial model and is less robust to measurement noise and physical assumptions in modeling. Compared to FWI, wavefield reconstruction inversion (WRI) is more robust to these uncertainties but faces high computational costs. To overcome these challenges, we develop a new form of WRI. This reformulation takes the form of a traditional FWI formula, which includes a medium-dependent weight function, and can be easily incorporated into the current FWI workflow. This weight function contains the covariance matrices to characterize the distribution of uncertainties in measurements and physical assumptions. We discuss various options of the theoretical covariance matrix of the new inversion method and show how they relate to various well-known approaches, including FWI, WRI, and extended FWI. On the basis of the above comparison, we propose a theoretical covariance matrix definition based on the source. Numerical experiments demonstrate that the proposed method with a source-dependent theoretical covariance matrix is more computationally efficient than conventional WRI, while preserving a certain degree of robustness.

INTRODUCTION

The ultimate goal of Full waveform inversion (Tarantola, 1984; Pratt, 1999) is to estimate subsurface medium parameters by minimizing the data discrepancy between simulations and observations. However, FWI is a non-linear and multi-modal problem that requires low-frequency and long-offset data to mitigate the local minima influence (Virieux and Operto, 2009; Plessix et al., 2010). That is to say, the success of FWI often relies on a good initial model. Otherwise, the data fitting error may be greater than half a wavelength, leading to the cycle-skipping problem (Beydoun and Tarantola, 1988; Pratt et al., 2008).

Many theories and methods have been proposed to alleviate the problems stated above. The multi-scale strategy in either the frequency domain (Bunks et al., 1995; Sirgue and Pratt, 2004) or the Laplace domain (Shin and Ho Cha, 2009) is commonly used to mitigate cycle-skipping and broaden the basin of attraction. Misfit functions in different forms or based on different variables (Luo and Schuster, 1991; Gee and Jordan, 1992; Ha et al., 2009; van Leeuwen and Mulder, 2010; Bozdag et al., 2011; Choi and Alkhalifah, 2015; Li et al., 2017; Lin et al., 2018) can also be used to alleviate the non-linearity issue. In the Adaptive waveform inversion (AWI, Warner and Guasch, 2016; Guasch et al., 2019), the data discrepancy between the predicted trace and the corresponding observed trace is measured using a Wiener filter to mitigate the influence of cycle-skipping. Apart from AWI, a new approach of matching the synthetic data and observed data using the Wasserstein distance (Metivier et al., 2017; Qiu et al., 2017; Yong et al., 2018) instead of the Euclidean distance is considered to be a promising method.

An alternative direction is the wavefield reconstruction inversion (WRI) method proposed by van Leeuwen and Herrmann (2013). By adding the wave equation as a penalty term to the objective function and solving the wavefield and model parameters alternately, WRI reduces the non-linearity of the inversion while expanding the optimization space. Furthermore, introducing a penalty term reduces the requirement for precise data fitting, making WRI less prone to cycle-skipping.

However, tuning the penalty parameter of the WRI is a challenging task. The basics rule to be followed is that a weak penalty should be chosen during the early iterations to facilitate data fitting, while the wave equation constraints should be progressively enforced during the later iterations. To address this issue, an automatic algorithm that balances the penalty at each iteration was proposed by da Silva (2015). Aghamiry et al. (2018) solves the penalty scalar problem by replacing the penalty method with an augmented Lagrangian method, where the Lagrange multipliers are updated with a classical dual ascent approach (IR-WRI as iteratively-refined WRI). Their experiments show that IR-WRI may significantly improve the convergence speed of WRI and make it easier to find the global minimum. However, solving an augmented wave equation is still inevitable in the IR-WRI, which is still computationally prohibited.

Another way to expand the space is the extended full-waveform inversion (EFWI, Symes, 2008; Abubakar et al., 2008), which enlarges the space by introducing non-physical degrees of freedom into the inversion. These additional degrees of freedom are not part of the fundamental physics used in the inversion and need to be suppressed during the inversion process. The dimension extended can be performed in a model space extension (Fu and Symes, 2017b) or in a source space extension (Huang et al., 2017). Fu and Symes (2017a) propose an adaptive multi-scale algorithm to reduce the computational cost of subsurface-offset EFWI (one kind of EFWI in the model space, Fu and Symes, 2017b) caused by the use of nonlocal wave physics (Mulder, 2013). In addition to the model-space EFWI, the EFWI in the source space (SEFWI) was also proposed with efficiency and less sensitivity of the cycle-skipping (Huang et al., 2018; Symes, 2020).

Despite progresses made in EFWI and WRI, they are still hampered by the huge computational overhead. Specifically, the extended wave equation introduces additional dimensions involving an integral over subsurface offset axis to compute the strain from stress, making subsurface offset extended FWI computationally infeasible. In comparison, WRI must solve an augmented wave equation, which involves the inverse computation of a complex matrix with broader bandwidth and larger condition numbers compared to the Helmholtz matrix.

This paper first argues that the extended inversion method essentially suppresses unknown measurement and theoretical uncertainties in practical cases, which is embodied in overcoming cycle-skipping caused by poor initial models. Then, we reformulated the WRI with these uncertainties as a residual-weighted FWI, where the weight term is a function of the measurement and theoretical covariance matrices. We discussed various theoretical covariance matrices and showed their connection with FWI, WRI, and SEFWI. Based on the above analysis, we propose a source-based theoretical covariance matrix definition, which simplifies the calculation of the weight function of the WRI in the form of FWI, making the new method more computationally efficient than the traditional WRI. Our main contribution are a theoretical framework that incorporates measurement and theoretical uncertainties and a definition of the theoretical covariance matrix for improving computational efficiency. Furthermore, the overall approach preserves the robustness of the WRI.

This paper is organized as follows. We first include the measurement and theoretical uncertainties in the conventional WRI formulation. Then, the WRI with measurement and theoretical uncertainties is reformulated into the form of the conventional FWI with a residual-weighted term. We discussed various options for the theoretical covariance matrices and connected them to FWI, WRI, and SEFWI. In particular, a rank-1 theoretical covariance matrix is proposed to reduce the computational cost of the WRI. Finally, numerical examples are implemented to examine the efficiency and performance of the proposed method.

THEORY

Wavefield reconstruction inversion

The forward modeling of seismic wave propagation can be described by the following wave equation:

$$\mathbf{A}(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i \, , \tag{1}$$

in which i is the source number index, $\mathbf{m}$ is a vector of the model parameters, $\mathbf{u}_i$ is the corresponding wavefield, $\mathbf{q}_i$ denotes the source term, and $\mathbf{A(m)}$ represents a discretized operator for the wave equation.

Uncertainties in data acquisition and forward simulation are the main instability factors when applying FWI to real data (Symes and Carazzone, 1991; Symes, 2008; Weglein, 2013; Brittan and Jones, 2019). Specifically, the measurement uncertainty is mainly generated by unpredictable factors in the recording environment, while the theoretical uncertainty is likely caused by imprecise physical assumptions of the Earth, incorrect wavelet estimation, and numerical errors in simulation. By considering such uncertainties, the observed data is expressed as:

$$\mathbf{d}_i = \mathbf{P}_i \mathbf{A}(\mathbf{m})^{-1}(\mathbf{q}_i + \boldsymbol{\epsilon}_p) + \boldsymbol{\epsilon}_m , \tag{2}$$

where $\mathbf{d}$ denotes the observed data, and $\mathbf{P}_i$ is the sampling operator that extracts data at the receivers excited by the ith source~~, $\mathbf{q}$ denotes the real source~~. $\boldsymbol{\epsilon}_p = \mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{q}$, and $\boldsymbol{\epsilon}_m = \mathbf{d} - \mathbf{Pu}$ are used to describe the theoretical uncertainty and the measurement uncertainty, respectively. Note that $\boldsymbol{\epsilon}_p$ is a variable associated with inaccurate laws of physics. Although $\boldsymbol{\epsilon}_p$ is placed in the source term for convenience and simplicity, we emphasize that $\boldsymbol{\epsilon}_p$ represents the influence caused by a combination of many factors, including but not limited to imprecise physical assumptions of the Earth, numerical errors in discretization, and incorrect wavelet estimation. Furthermore, it should be noted that equation 2 is an exact symbolic equation describing wave propagation and observations in real scenarios, while equation 1 is a theoretically derived equation used for simulation guided by certain physical assumptions of the Earth. One can only use equation 1 for seismic imaging and inversion since $\boldsymbol{\epsilon}_p$ and $\boldsymbol{\epsilon}_m$ are unpredictable in real scenarios. However, in this paper, existing methods are generalized and optimized by acknowledging the existence of measurement and theoretical uncertainties and describing their distribution. To simplify the derivation and calculation, a simple assumption is made to these uncertainties, which all satisfy the Gaussian distribution, delineated as:

$$\boldsymbol{\epsilon}_m \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{m}, \Sigma_m) , \tag{3}$$

$$\boldsymbol{\epsilon}_p \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{p}, \Sigma_p) , \tag{4}$$

where $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ represents a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance Σ. We assume that $\boldsymbol{\mu}_\mathbf{m} = \boldsymbol{\mu}_\mathbf{p} = 0$ for the sake of simplicity. Other non-zero means can also be included in our framework.

The assumptions of equations 3 and 4 lead to a weighted least-squares problem (see Appendix A for details):

$$\min_{\mathbf{u},\mathbf{m}} \sum_i \|\mathbf{P}_i\mathbf{u}_i - \mathbf{d}_i\|^2_{\Sigma_m} + \|\mathbf{A}(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i\|^2_{\Sigma_p}, \quad (5)$$

in which $\|\mathbf{r}\|^2_\mathbf{W} = \mathbf{r}^*\mathbf{W}^{-1}\mathbf{r}$ and * denotes the adjoint of $\mathbf{r}$. If both covariance matrices are a multiple of the identity diagonal matrix, i.e., $\Sigma_p = \sigma_p\mathbf{I}, \Sigma_m = \sigma_m\mathbf{I}$, we can obtain another form of the traditional WRI based on the penalty parameter $\lambda = \sigma_p^{-1} / \sigma_m^{-1}$. An alternating optimization method (Bezdek and Hathaway, 2002) and a variable projection method (Golub and Pereyra, 2003) are commonly used to solve equation 5. A two-step procedure is usually adopted in the alternating optimization algorithm. In the first step, equation 5 is minimized over the wavefield $\mathbf{u}$ for fixed model parameters $\mathbf{m}$ by solving:

$$(\mathbf{P}_i^*\Sigma_m^{-1}\mathbf{P}_i + \mathbf{A}^*\Sigma_p^{-1}\mathbf{A})\hat{\mathbf{u}}_i = (\mathbf{P}_i^*\Sigma_m^{-1}\mathbf{d}_i + \mathbf{A}^*\Sigma_p^{-1}\mathbf{q}_i). \quad (6)$$

In the second step, the model is updated using a gradient step with step-length $\alpha > 0$; $\mathbf{m} := \mathbf{m} - \alpha\mathbf{g}$ with:

$$\mathbf{g} = \sum_i G(\mathbf{m}, \hat{\mathbf{u}}_i)^* \Sigma_p^{-1}(\mathbf{A}(\mathbf{m})\hat{\mathbf{u}}_i - \mathbf{q}_i), \quad (7)$$

in which $G(\mathbf{m},\hat{\mathbf{u}})$ is the Jacobian matrix of $\mathbf{A}(\mathbf{m})\hat{\mathbf{u}}$. Though an exact solution for the model parameter can be obtained (van Leeuwen and Herrmann, 2013; Aghamiry et al., 2019), the gradient descent method is chosen in this paper, we can directly calculate the model domain Hessian (van Leeuwen and Herrmann, 2015) to further improve the inversion result.

Equation 6 is the augmented wave equation, which replaces forward modeling and adjoint state calculations, thereby expanding the search space and avoiding the storage of synthetic data for all sources compared to the classical solutions for FWI. However, the memory requirements of a traditional WRI can become very large, as the matrix bandwidth of $\mathbf{P}^*\mathbf{P}+\mathbf{A}^*\mathbf{A}$ and the condition number significantly increases with the model size.

In summary, by including measurement and theoretical uncertainties in the inversion, WRI can reduce the non-linearity of the inversion problem. However, how to best determine the theoretical covariance matrix and reduce the expensive computational cost of equation 6 remain challenges for WRI.

An alternative formulation of WRI

In this section, an alternative formulation for WRI with measurement and theoretical uncertainties is derived. This reformulation takes the form of conventional FWI with a medium-dependent residual weight (WRI in the FWI form, FWRI). First, based on the WRI method, considering measurement and theoretical uncertainties, the formulations of SEFWI can be obtained by introducing a new source-like variable:

$$\mathbf{r}_i = \mathbf{A}(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i \,. \tag{8}$$

Note that $\mathbf{r}$ is not the same as $\boldsymbol{\epsilon}_p$. $\mathbf{r}$ represents the extended source mainly used for the following derivation, and $\boldsymbol{\epsilon}_p$ is the existing unknown variable. The inversion problem of equation 5 becomes (van Leeuwen, 2019):

$$\min_{\mathbf{r},\mathbf{m}} \sum_i \left\| \mathbf{P}_i \mathbf{A}(\mathbf{m})^{-1}(\mathbf{r}_i + \mathbf{q}_i) - \mathbf{d}_i \right\|_{\Sigma_m}^2 + \left\| \mathbf{r}_i \right\|_{\Sigma_p}^2 . \quad (9)$$

Equation 9 brings us a new relation between WRI and EFWI: different choices for Σ_p lead to other kinds of extended FWI methods. For example, if Σ_p^{-1} becomes a function of the source distance in the frequency domain, we will have the objective function of the source extended FWI (Equation 10 in Huang et al., 2018). Furthermore, the connection of the above two methods above will help us define the theoretical covariance matrix. More details will be discussed in the following section. Note that Equations 8 and 9 are derived from measurement and theoretical uncertainties containing probability-related variables (covariance matrices), which are slightly different from Wang et al. (2016).

Through algebra (see Appendix B for more details), we can show that equation 9 is equivalent to an objective function in the traditional FWI form with a medium-dependent weight (van Leeuwen, 2019):

$$\phi(\mathbf{m}) = \sum_i \left\| \mathbf{P}_i \mathbf{A}(\mathbf{m})^{-1} \mathbf{q}_i - \mathbf{d}_i \right\|_{\Sigma_i(\mathbf{m}) + \Sigma_m}^2 , \quad (10)$$

with

$$\Sigma_i(\mathbf{m}) = \mathbf{P}_i (\mathbf{A}(\mathbf{m})^* \Sigma_p^{-1} \mathbf{A}(\mathbf{m}))^{-1} \mathbf{P}_i^* . \quad (11)$$

Compared to equation 5, the reformulated expressions (equations 10 and 11), referring to as the FWRI, indicate that the influences of measurement and theoretical uncertainties for inversion can be considered as the weight function or the data-domain Hessian matrix in the traditional FWI (Gholami et al., 2021b). However, both the measurement and theoretical uncertainties ~~satisfy the identity~~ are usually described by the diagonal covariance matrices in existing methods. In addition, the new formula of WRI shows another strategy for applying WRI in the time domain by avoiding the calculation of augmented wave equations. The large-scale data domain Hessian matrix in the time domain FWRI can be computed by some fast approximation algorithms (Yuzhao et al., 2022).

Remarkably, the gradient of the new objective function with respect to the model parameters has a simple expression, similar to that of conventional FWI:

$$\nabla\phi(\mathbf{m}) = \sum_i -2\mathbf{u}_{0_i}^{*}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{v}_{0_i} + 2\mathbf{v}_{0_i}^{*}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{w}_{0_i}, \qquad (12)$$

in which:

$$\mathbf{u}_{0_i} = \mathbf{A}^{-1}\mathbf{q}_i, \qquad (13)$$

$$\mathbf{w}_{0_i} = \mathbf{A}^{-1}\Sigma_p\mathbf{v}_{0_i}, \qquad (14)$$

$$\mathbf{v}_{0_i} = (\mathbf{P}_i\mathbf{A}^{-1})^{*}(\Sigma(\mathbf{m})_i + \Sigma_m)^{-1}(\mathbf{P}_i\mathbf{A}(\mathbf{m})^{-1}\mathbf{q}_i - \mathbf{d}_i) = (\mathbf{P}_i\mathbf{A}^{-1})^{*}\hat{\mathbf{h}}_i, \qquad (15)$$

where $\hat{\mathbf{h}} = (\Sigma(\mathbf{m})_i + \Sigma_m)^{-1}\mathbf{h}$ is the weighted residual and $\mathbf{h} = \mathbf{P}_i\mathbf{A}(\mathbf{m})^{-1}\mathbf{q}_i - \mathbf{d}_i$ is the classical FWI residual. Both terms in gradient of the FWRI are susceptible to cycle-skipping issue and robust to non-physical data due to the weight residual (Yuzhao et al., 2022). Specifically, the second term in equation 10 appears as the wave path from the receiver to the reflector, which provides lowwavenumber model updates. Moreover, the FWRI can be readily adapted to the standard FWI workflow. We will introduce the fully detailed FWRI algorithm after discussing the definition of the theoretical covariance matrix and the computation of the weighted residual.

Theoretical covariance matrix

To accurately solve equation 10, we first need to define two covariance matrices that describe theoretical and measurement uncertainties, respectively. Since the measurement uncertainty is caused by unpredictable environment factors, which is highly random in nature, it can be represented by the identity covariance matrix $\Sigma_m = \sigma_m \mathbf{I}$. Therefore, in the following, we concentrate on how to define the theoretical covariance matrix.

In principle, the definition of the theoretical covariance matrix is arbitrary since statistical testing of seismic inversion is impractical. However, for the following reasons, we can provide a reasonable definition that maintains the advantages of WRI while being computationally efficient.

First, the theoretical covariance matrix is utilized to evaluate the uncertainties of the wave equation we used in the inversion compared to the actual situation. Since no statistical definition can be given, choosing known variables as the theoretical covariance matrix is relatively appropriate. Furthermore, the analysis of equation 9, namely the derivation of the SEFWI by defining a function related to the source (source distance annihilator), also indicates that the related variables of the source, i.e., source location and source frequency spectrum, can be used as the definition of the theoretical covariance matrix.

Furthermore, we can divide the weight function by defining the theoretical covariance matrix as a random variable and its transpose:

$$\Sigma_i(\mathbf{m}) = (\mathbf{P}_i\mathbf{A}(\mathbf{m})^{-1}\mathbf{b})(\mathbf{P}_i\mathbf{A}(\mathbf{m})^{-1}\mathbf{b})^*, \tag{16}$$

where

$$\Sigma_p = \mathbf{bb}^*. \tag{17}$$

It can be seen that the weight function consists of forward modeling excited by $\mathbf{b}$. Intuitively, letting $\mathbf{b} = \mathbf{q}$ allows us to dispense with additional forward modeling of the weight function.

Therefore, we propose a theoretical covariance matrix definition from a limited selection based on the above analysis: $\Sigma_p = \mathbf{q}_i\mathbf{q}_i^*$. Note that the theoretical covariance matrix defined is not invertible. However, we are solving equation 10, and the proposed method still holds as long as the weight function is invertible. Furthermore, we demonstrate that this definition is similar in some respects to other definitions and also brings computational advantages. However, we acknowledge that other forms of the theoretical covariance matrix may existed and provide same benefits. Note that our definition of the theoretical covariance matrix does not jeopardize the correlations that form the data domain Hessian, since we are only replacing the source of the Green function or forward modeling to avoid additional modeling calculations.

To compare the theoretical covariance matrices and the weight function, a numerical experiment is carried out to illustrate their differences and connections based on different definitions. The test model is a constant velocity model of $210\,\mathrm{m} \times 210\,\mathrm{m}$ with 10m spatial sampling. The source is located at a depth of 20m and horizontally at 100m. The theoretical covariance matrix, noise distributions, and weight functions are plotted in Figures 1, 2, and 3, respectively.

Figure 1a shows the diagonal matrix $\Sigma_p = \sigma \mathbf{I}$, the theoretical covariance matrix definition in traditional WRI. The corresponding theoretical noise distribution (Figure 2a) is random. Figure 1b shows the proposed theoretical covariance matrix $\Sigma_p = \mathbf{q}\mathbf{q}^*$, which contains only one non-zero element at the source location, consistent with theoretical noise (Figure 2b). Figure 1c shows the source distance annihilator $\left(\Sigma_p\right)_{i,i} = 1/(\| \mathbf{x}_i - \mathbf{x}_s \|_2^2)$ proposed by Huang et al. (2018). The corresponding theoretical noise in Figure 2c decreases as the distance between the source location and the model point increases. As shown in Figures 1 and 2, the theoretical noise associated with the theoretical covariance matrix in the traditional WRI is randomly distributed, while the theoretical noise and the theoretical covariance matrix based on the other two definitions are highly correlated. By comparing the theoretical covariance matrix and the corresponding noise, we can presume that our definition is a limiting case of the source distance annihilator definition; that is, only the value at the true source location is considered.

Next, we show the weight functions corresponding to different definitions in Figure 3 to demonstrate how different definitions directly affect the inversion. The weight function of traditional WRI (Figure 3a) and SEFWI (Figure 3c) are similar, and the energy is uniformly diffused from the center to the outside. The weight function corresponding to the definition we proposed is not uniform. Although it also radiates from the center, the energy of the center point (source location) is still more vital than that of other points. Moreover, this feature is consistent with the above conclusion that $\Sigma_p = \mathbf{q}_i\mathbf{q}_i^*$ is an extreme case of a source-distance annihilator. Furthermore, Figure 3 demonstrates the similarity between the conventional WRI and SEFWI in a novel way by comparing the weight functions.

Note that in SEFWI, the source distance annihilator is used to shrink the extended source. In our framework, the source distance annihilator is an invariant variable (theoretical variance matrix) used to describe the theoretical uncertainty. It only exhibits model-dependent characteristics when operating on the extended source, while the proposed covariance matrix definition is the same as the source distance annihilator. Moreover, we have only discussed the definition of the theoretical covariance matrix; the solutions for WRI, FWRI, and SEFWI are still different. Nonetheless, according to the above definitions, they can be solved under the framework presented in our previous section, albeit at a relatively high computational cost. For example, the computation of SEFWI would involve forward modeling of a vector with elements $s_i = 1/\|\mathbf{x}_i - \mathbf{x}_s\|_2$.

In conclusion, we show that by giving different theoretical covariance matrix definitions, different inversion methods can be derived from equation 9. The theoretical analysis and numerical comparisons show that our definition of the theoretical covariance matrix that is more computationally efficient than others. Our definition of the theoretical covariance matrix in terms of the source-dependent variables can also be cross-verified by the WRI mechanism (Gholami et al., 2021a), which states that the kinematic errors generated during the wavefield reconstruction can be compensated by the extended sources. It should be noted that, in addition to our definition, other types of source-related variables can also be used to define the theoretical covariance matrix and thus develop other inversion algorithms.

Algorithm

In addition to the theoretical covariance matrix, the calculation of weighted residual must be solved before introducing the proposed FWRI algorithm with the $\mathbf{qq}^*$ definition. As mentioned previously, the computation of the weighted residual is a multidimensional deconvolution problem, and the weight function is the data domain Hessian when the theoretical covariance matrix defaults to the unit diagonal matrix. Approximations are widely used to in extended WRI methods involving source and data inversion to avoid direct computation of the data domain Hessian, due to its large scale (Aghamiry et al., 2019; Gholami et al., 2021b). For example, in Gholami et al. (2021b), the data domain Hessian was approximated as an update step in the data inversion step to facilitate computation.

In the frequency domain, one can choose the LSQR method or gradient descent method to calculate the weighted residual. However, as with most computations involving the Hessian, the accuracy of the weighted residual cannot be guaranteed due to the poor conditioning of the Hessian. Here we let the sampled wavefield be written as:

$$\mathbf{u}_{\mathbf{p}i} = \mathbf{P}_i \mathbf{A}^{-1} \mathbf{q}_i . \tag{18}$$

According to our definition, the weight term can be written as:

$$(\Sigma_m + \Sigma(\mathbf{m})_i)^{-1} = (\sigma_m \mathbf{I} + \mathbf{u}_{\mathbf{p}i} \mathbf{u}_{\mathbf{p}i}^*)^{-1} . \tag{19}$$

Considering its unique form and using the Sherman-Morrison identity gives us:

$$(\Sigma_m + \Sigma(\mathbf{m})_i)^{-1} = \sigma_m^{-1} (\mathbf{I} + \frac{\mathbf{u}_{\mathbf{p}i} \mathbf{u}_{\mathbf{p}i}^* \sigma_m^{-1}}{1 + \mathbf{u}_{\mathbf{p}i}^* \sigma_m^{-1} \mathbf{u}_{\mathbf{p}i}}) . \tag{20}$$

Therefore, calculating the weighted residual $\hat{\mathbf{h}}$ becomes relatively easy and affordable: for each shot and frequency, the size of Equation 20 is the square of the number of receivers. Now we can employ Algorithm 1 to execute the proposed method. Compared with traditional FWI, FWRI requires limited additional computations to calculate residual and gradients.

To quantitively evaluate the computational efficiency of the proposed FWRI, numerical experiments are conducted using FWI, WRI, and the -based FWRI, respectively, with a Dell Vostro-5390 laptop equipped with i7-8565U and 8 GB RAM. All experiments are carried out using one source and twenty receivers for a single frequency.

The computational comparisons of the traditional FWI, WRI, the FWRI with the identity matrix, and the proposed -based FWRI are shown in Table 1. The standard FWI workflow involves forward modeling based on the current model, adjoint source, computation for the adjoint state, and gradient calculation. Moreover, compared to FWI, WRI does not need to calculate the adjoint source, and its adjoint state calculation is simple. In contrast, forward modeling in WRI is essential to solving an augmented wave equation, which requires considerable memory due to the wide bandwidth.

The calculation of the FWRI with the identity covariance matrix becomes much more complicated than that of the WRI and FWI: in the adjoint source calculations, the weight function and its inverse need to be calculated. The adjoint state requires additional forward modeling. In contrast, the computation of the $\mathbf{qq}^*$-based FWRI is much simpler since the Sherman–Morrison formula can be used to obtain adjoint source without additional forward modeling.

Moreover, the number of PDE that need to be solved using each method when calculating the model updates is also provided for comparison. For a more accurate comparison, we separate the solutions of the PDE (i.e., the standard wave equation) and the augmented wave equation

(AWE). The AWE solver requires more memory and is more time-consuming than the PDE solver due to the wider bandwidth. Moreover, the calculation of AWE is unstable due to the large condition number.

In this case, traditional FWI requires 2 PDE solvers, while traditional WRI requires 1 AWE solver. Due to the $\mathbf{A}^{-*}\mathbf{P}^{*}$ in $\mathbf{h}_{w}$, FWRI with an identity covariance matrix requires $3+N_{receiver}$ PDE solvers, where $N_{receiver}$ PDE solvers come from the explicit inversion of the matrix $\mathbf{A}^{-*}\mathbf{P}^{*}$. In contrast, $\mathbf{qq}^{*}$-based FWRI requires only 2 PDE solvers, meaning that the proposed method based on the given theoretical covariance matrix definition is equivalent to FWI in terms of computational cost.

Compared with other extended inversion methods, according to our theoretical covariance matrix description, SEFWI requires an additional PDE solution or forward modeling excited by $s_i = 1/\| \mathbf{x}_i - \mathbf{x}_s \|_2$. In addition, a WRI-dual method proposed by Rizzuti et al., (2019, 2021) requires 4 PDE solvers.

As mentioned above, PDE solvers and AWE solvers have different computational costs and memory requirements that are related to the computational area or model size. To visualize the computational cost of four methods, we plot their computational time for a single iteration at different model sizes in Figure 4. The test is based on a constant model of square matrix size. The X labels in Figure 4 indicate the number of grid points in one direction.

With the increase in grid points, the time required to complete WRI and identity matrix-based FWRI grows very fast, while FWI and $\mathbf{qq}^{*}$-based FWRI show relatively slow growth, indicating that the former methods are more time-consuming. In particular, the solution time of identity matrix-based FWRI is the most time-consuming due to its $N_{receiver}$ PDE solver requirement, as described above. However, we draw this curve for reference only, as it is not the priority of this test. The two nearly identical curves representing FWI and $\mathbf{qq}^*$-based FWRI also support our above claim that $\mathbf{qq}^*$-based FWRI is computationally equivalent to FWI.

NUMERICAL TESTS

The wave equation is discretized by a 9-point discretization of the 2D Helmholtz operator with an absorbing condition, and the optimization problem is solved by the L-BFGS (Fabien et al., 2017) iteration method.

Non-linearity

This test supports our claim that the weight function-based FWRI is less prone to falling into a local minimum by incorporating uncertainties into the inversion. That is, the effect of the theoretical uncertainties we introduced above on the inversion is mainly manifested as cycle-skipping caused by poor initial models. As previously mentioned, defining the theoretical covariance matrix as a unit diagonal matrix will lead to traditional WRI, and van Leeuwen and Herrmann (2013, 2015) have discussed its reduced non-linearity. Here, we will investigate the non-linearity of FWRI with a defined theoretical covariance matrix $\mathbf{qq}^*$.

More specifically, a constant velocity is used to simulate data which serves as input to the objective function calculation ~~a constant velocity is considered the input of misfit functions~~, and the actual velocity is $c = 2000\,\mathrm{m/s}$. A single source of $5\,\mathrm{Hz}$, and 201 receivers, both at a depth of 25m, is used for the test. We scan the velocity from $1750\,\mathrm{m/s}$ to $2250\,\mathrm{m/s}$ to calculate the objective function.

Figure 5 show the comparison of misfit functions. It clearly shows the reduced non-linearity of the proposed method: the objective function with the weight function (red line) has no local minimum in the velocity range considered. It directly converges to the actual global minimum, while the objective function of the traditional FWI (blue line) presents three secondary minima and a global minimum. We also plot the SEFWI result (green line) because it is only different from the FWRI in the covariance matrix. Numerical experiments show that it also has good convexity.

Camembert model

To demonstrate the robustness of the FWRI, we apply the proposed method to the Camembert model (Gauthier et al., 1986), which includes circular anomaly (Figure 6a). The model contains 101×101 grid points, and the grid spacing in both directions is 10 m. The velocity of the anomaly is $2600\,\mathrm{m/s}$, and the homogeneous background is $2000\,\mathrm{m/s}$.

We use transmissions to perform the inversion; in this case, we will have sufficient information to recover the large-scale anomaly. Therefore, the acquisition consists of 9 sources evenly distributed on the surface and 50 receivers located at the bottom. The inversion starts from the homogeneous background model, performed by frequency slices, and the corresponding inversion frequency band is $[4,12]\,\mathrm{Hz}$ with an interval of $1\,\mathrm{Hz}$.

The inversion models obtained using FWI, WRI, and FWRI are shown in Figures 6b, 6c, and 6d, respectively. The WRI method is basically the same as the FWRI method but has a different covariance matrix. Considering the computational cost, we apply conventional WRI in the following tests rather than the FWRI with an identity theoretical covariance matrix. The FWI result is far from the accurate model; only the top and bottom of the abnormal body that can generate the waveform are accurately restored. Moreover, part of the low-frequency models is also recovered due to the transmissions. The central part of the anomaly is hardly inverted. However, the WRI and FWRI methods provide accurate results, and both models are close to the ground truth and accurately invert the large anomaly.

BP model

Finally, we evaluate the $\mathbf{qq}^*$-based FWRI method with part of the BP model (Figure 7a; Billette and Brandsberg, 2005) based on a flawed initial model (Figure 7b). We use a Gaussian filter with vertical and horizontal lengths of 1500 m to smooth the real model to obtain the initial model. The size of the selected BP model is 160×387, with a grid interval of 25m. A total of 38 shots are simulated on the surface, with receivers evenly distributed in 250 m intervals at the surface with a maximum offset of 9675m. The inversion is performed frequency-by-frequency, and the frequency band is $[3,12]\,\mathrm{Hz}$ with a $0.5\,\mathrm{Hz}$ frequency interval.

In addition, we add very little Gaussian noise (Signal-to-noise ratio, SNR=0.5) with the same covariance matrix for all frequency data to emphasize model or theoretical uncertainties. Traditional FWI, WRI, and $\mathbf{qq}^*$-based FWRI are performed for comparison.

The inverted velocities are shown in Figure 8. Figure 8a is the FWI result. Due to the wrong initial velocity, the inverted velocity of FWI contains some artifacts and is far away from the true model. Figure 8c shows the FWRI method with defined covariance matrix $\mathbf{qq}^*$, which slightly differs from the WRI result (Figure 8b). We can see that both methods accurately recovered the salt body, which demonstrates that the proposed approach and the well-defined theoretical covariance matrix can handle an inversion problem with strong non-linearity. However, the subsalt structure on the left cannot be recovered accurately from these two methods due to the limited illumination caused by the salt structure. However, considering the computational cost, FWRI is an excellent alternative to WRI.

DISCUSSION

It is noted that measurement and theoretical uncertainties can reduce the non-linearity of the inversion through the weight term or Hessian matrix. Since there is no need to calculate the augmented wave equation, FWRI can be applied to the time domain, and the influence of theoretical uncertainties on the inversion can be further intuitively analyzed by using variables such as the time domain wavefield. However, due to some computational issues related to the weight function, the proposed method cannot currently be used efficiently in the time domain. Specifically, the scale of the data domain Hessian matrix in the time domain is $Nt \times Nr \times Nt \times Nr$ (Nt represents the travel time, and Nr is the receiver number). This large-scale characteristic of the Hessian makes it challenging to perform calculations in the time domain directly. The

Sherman-Morrison identities used to compute the weight function hold theoretically in both the time and frequency domains. Due to the large scale ($Nt \times Nr \times Nt \times Nr$) of $\mathbf{u}_{\mathbf{p}i}\mathbf{u}_{\mathbf{p}i}^{*}$ or equation 20, calculating the weight function in the time domain requires a large amount of memory or a significant ~~lengthy~~ computational time, even from the perspective of the total calculation of FWI. A cost-efficient time-domain solution for WRI and its weighted residual is under study.

Furthermore, as shown by Louboutin et al. (2020), traditional WRI can provide reliable inversion results when inaccurate forward operators are used. The theoretical analysis shows that the proposed method and the definition of theoretical variance can maintain this robustness. More specifically, when we adopt the definition of $\mathbf{qq}^{*}$, the data domain Hessian matrix is essentially a matrix composed of $\mathbf{Pu}(\mathbf{Pu})^{*}$, and the calculation of the weighted residual becomes $\mathbf{Pu}(\mathbf{Pu})^{*}\hat{\mathbf{h}} = \mathbf{h}$. When the original residual contains other responses that the wave equation cannot describe, the calculation of the weight function and weighted residual will reduce the weight of these data. In this case, subsequent gradient computations are less affected by non-physical data.

CONCLUSION

We have derived an alternative formula for extended full-waveform inversion from wavefield reconstruction inversion with measurement and theoretical uncertainties. The derivation of the new formula shows that different theoretical covariance matrix definitions in WRI correspond to different extended FWI methods (e. g., source extended FWI). By choosing the theoretical covariance matrix as low rank, we obtain a computationally feasible method for the extended FWI. Several experiments have been described to illustrate the effectiveness of the proposed method. Testing the Camembert model emphasizes the validity of the defined theoretical covariance matrix. BP model tests show that the new method preserves the robustness of WRI and has the ability to handle non-linear problems.

Future research will focus on searching for more effective choices for the theoretical covariance matrix. Although the approximation of the forward operator indicates a method of estimating the covariance matrix, it is necessary to further study the choice of different theoretical covariance matrices and their influence on the inversion.

APPENDIX A

In this section, we derive equation 5 from the work of Tarantola (2005). The shot index i was omitted for ease of notation.

Define the joint probability density Θ describing the correlations that correspond to our physical theory, together with the inherent uncertainties of the theory as:

$$\Theta(\mathbf{u},\mathbf{m}) = \theta(\mathbf{u}\,|\,\mathbf{m})\mu_{\mathbf{M}}(\mathbf{m}), \tag{A-1}$$

where $\mathbf{m}$ is the model parameters, $\mathbf{u}$ denotes the wavefield, θ is the conditional probability density, and $\mu_{\mathbf{M}}$ represents the prior information of the model parameters.

Assuming the prior information of the model is constant and the conditional probability density is a Gaussian probability density with $\boldsymbol{\mu}_{\mathbf{p}} = 0$:

$$\Theta(\mathbf{u}\,|\,\mathbf{m}) = \mathbf{exp}(-\frac{1}{2}(\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{q})^{*}\Sigma_{p}^{-1}(\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{q})). \tag{A-2}$$

The measurement uncertainties can be described by a Gaussian probability density $\rho_{\mathbf{D}}(\mathbf{d}) = \rho_{\mathbf{D}}(\mathbf{Pu})$ centered at $\mathbf{d}$:

$$\rho_{\mathbf{D}}(\mathbf{d}) = \mathbf{exp}(-\frac{1}{2}(\mathbf{Pu}-\mathbf{d})^{*}\Sigma_{m}^{-1}(\mathbf{Pu}-\mathbf{d})). \tag{A-3}$$

These two states of information are combined to produce the posterior state of information:

$$\Psi(\mathbf{d},\mathbf{m}) = \Theta(\mathbf{u}\,|\,\mathbf{m})\rho_{\mathbf{D}}(\mathbf{d}). \tag{A-4}$$

The posterior information in the model space is given by the marginal probability density:

$$\Psi_{\mathbf{M}}(\mathbf{m}) = \int_{\mathbf{D}} \Theta(\mathbf{u}\,|\,\mathbf{m})\rho_{\mathbf{D}}(\mathbf{d})d\mathbf{d}, \tag{A-5}$$

which gives a maximum a posteriori (MAP) problem:

$$\Psi_{\mathbf{M}}(\mathbf{m}) = \mathbf{exp}(-\frac{1}{2}(\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{q})^{*}\Sigma_{p}^{-1}(\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{q}))\mathbf{exp}(-\frac{1}{2}(\mathbf{Pu}-\mathbf{d})^{*}\Sigma_{m}^{-1}(\mathbf{Pu}-\mathbf{d})), \tag{A-6}$$

which is equivalent to equation 5.

APPENDIX B

This appendix provides mathematical proof of FWI formed WRI. To derive equations 10 and 11, we need to define a new wavefield $\mathbf{f}$:

$$\mathbf{f} = \mathbf{A}^{-1}\mathbf{r}, \tag{B-1}$$

where $\mathbf{A}$ is the forward operator and $\mathbf{r}$ is the source defined by equation 8. Using $\mathbf{h}$ to represent the residual:

$$\mathbf{h} = \mathbf{d} - \mathbf{PA}^{-1}\mathbf{q}, \tag{B-2}$$

where $\mathbf{d}$ is the observed data, $\mathbf{P}$ is the sampling operator, and $\mathbf{q}$ denotes the source used for modeling and inversion. Note that we omit the shot index i ~~for~~ to simplify notion. The objective function equation 9 becomes:

$$\min \left\|\mathbf{Pf} - \mathbf{h}\right\|_{\Sigma_m}^{2} + \left\|\mathbf{Af}\right\|_{\Sigma_p}^{2}, \tag{B-3}$$

with f defined as:

$$\mathbf{f} = (\mathbf{P}^*\Sigma_m^{-1}\mathbf{P} + \mathbf{A}^*\Sigma_p^{-1}\mathbf{A})\mathbf{P}^*\Sigma_m^{-1}\mathbf{h}, \tag{B-4}$$

$$\mathbf{f} = (\mathbf{A}^*\Sigma_p^{-1}\mathbf{A})^{-1}\mathbf{P}^*(\mathbf{P}^*(\mathbf{A}^*\Sigma_p^{-1}\mathbf{A})^{-1}\mathbf{P} + \Sigma_m)^{-1}\mathbf{h}. \tag{B-5}$$

The first definition can be obtained directly from equation B-3, while the second definition follows from the matrix identity (6.525) in Tarantola (2005).

Thus, we can write:

$$\mathbf{Pf} = \Sigma(\mathbf{m})(\Sigma(\mathbf{m}) + \Sigma_m)^{-1}\mathbf{h}, \tag{B-6}$$

$$\mathbf{Af} = \Sigma_p A^{-*}\mathbf{P}^*(\Sigma(\mathbf{m}) + \Sigma_m)^{-1}\mathbf{h}, \tag{B-7}$$

with $\Sigma(\mathbf{m})$ defined in equation 11. This yields:

$$\|\mathbf{Af}\|_{\Sigma_p}^2 = \mathbf{h}^*(\Sigma(\mathbf{m}) + \Sigma_m)^{-1}\Sigma(\mathbf{m})(\Sigma(\mathbf{m}) + \Sigma_m)^{-1}\mathbf{h}, \tag{B-8}$$

$$\|\mathbf{Pf} - \mathbf{h}\|_{\Sigma_m}^2 = \mathbf{h}^*(\Sigma(\mathbf{m})(\Sigma(\mathbf{m}) + \Sigma_m)^{-1} - \mathbf{I})^*\Sigma_m^{-1}(\Sigma(\mathbf{m})(\Sigma(\mathbf{m}) + \Sigma_m)^{-1} - \mathbf{I})\mathbf{h}. \tag{B-9}$$

Assembling all the terms and factoring out $(\Sigma(\mathbf{m}) + \Sigma_m)^{-1}$, we obtain:

$$\phi(\mathbf{m}) = \mathbf{h}^*(\Sigma(\mathbf{m}) + \Sigma_m)^{-1}(\Sigma(\mathbf{m}) + \Sigma_m)(\Sigma(\mathbf{m}) + \Sigma_m)^{-1}\mathbf{h}. \tag{B-10}$$

The above equation reduces to

$$\phi(\mathbf{m}) = \mathbf{h}^*(\Sigma(\mathbf{m}) + \Sigma_m)^{-1}\mathbf{h}, \tag{B-11}$$

which is our result. Note that there is no stand-alone inverse operation for the projector $\mathbf{P}$ or Σ_p, which are always calculated in the form $\mathbf{PA}^{-1}\mathbf{s}$ or its adjoint $(\mathbf{PA}^{-1}\mathbf{s})^*$.

APPENDIX C

The gradient can be expressed as:

$$\mathbf{g}(\mathbf{m}) = \sum_i \frac{\partial \phi(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial}{\partial \mathbf{m}} (\mathbf{P}_i \mathbf{A}(\mathbf{m})^{-1} \mathbf{q}_i - \mathbf{d}_i)^* (\Sigma(\mathbf{m})_i + \Sigma_m)^{-1} (\mathbf{P}_i \mathbf{A}(\mathbf{m})^{-1} \mathbf{q}_i - \mathbf{d}_i), \quad \text{(C-1)}$$

where $\mathbf{g}$ is the gradient, ϕ is the objective function, $\mathbf{m}$ denotes the model parameters, and $\mathbf{P}$ is the sampling operator, $\mathbf{A}$ denotes the forward operator, $\mathbf{q}$ is the source, and $\mathbf{d}$ is the observed data. $(\Sigma(\mathbf{m})_i + \Sigma_m)$ is the weight function as defined in equation 11.

This expression may be derived using the quotient-rule for matrix-differentiation:

$$\partial \mathbf{A}^{-1} = -\mathbf{A}^{-1} \partial \mathbf{A} \mathbf{A}^{-1}. \quad \text{(C-2)}$$

For the first term, we have:

$$\frac{\partial}{\partial \mathbf{m}} (\mathbf{P}_i \mathbf{A}^{-1} \mathbf{q}_i - \mathbf{d}_i)^* = -(\mathbf{P}_i \mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{A}^{-1} \mathbf{q}_i)^* = -(\mathbf{A}^{-1} \mathbf{q}_i)^* \frac{\partial \mathbf{A}}{\partial \mathbf{m}} (\mathbf{P}_i A^{-1})^*. \quad \text{(C-3)}$$

For the second term, we have:

$$(\mathbf{P}_i \mathbf{A}^{-1} \mathbf{q}_i - \mathbf{d}_i)^* \frac{\partial}{\partial \mathbf{m}} (\Sigma(\mathbf{m})_i + \Sigma_m)^{-1} (\mathbf{P}_i \mathbf{A}^{-1} \mathbf{q}_i - \mathbf{d}_i), \quad \text{(C-4)}$$

in which

$$\frac{\partial}{\partial \mathbf{m}}(\Sigma(\mathbf{m})_i+\Sigma_m)^{-1}=-(\Sigma(\mathbf{m})_i+\Sigma_m)^{-1}\frac{\partial \Sigma(\mathbf{m})_i}{\partial \mathbf{m}}(\Sigma(\mathbf{m})_i+\Sigma_m)^{-1}, \quad \text{(C-5)}$$

$$\frac{\partial \Sigma(\mathbf{m})_i}{\partial \mathbf{m}}=\frac{\partial}{\partial \mathbf{m}}\mathbf{P}_i(\mathbf{A}^*\Sigma_p^{-1}\mathbf{A})^{-1}\mathbf{P}_i^*=\mathbf{P}_i\frac{\partial}{\partial \mathbf{m}}(\mathbf{A}^*\Sigma_p^{-1}\mathbf{A})^{-1}\mathbf{P}_i^*, \quad \text{(C-6)}$$

and

$$\frac{\partial}{\partial \mathbf{m}}(\mathbf{A}^*\Sigma_p^{-1}\mathbf{A})^{-1}=-(\mathbf{A}^*\Sigma_p^{-1}\mathbf{A})^{-1}\frac{\partial}{\partial \mathbf{m}}(\mathbf{A}^*\Sigma_p^{-1}\mathbf{A})(\mathbf{A}^*\Sigma_p^{-1}\mathbf{A})^{-1}. \quad \text{(C-7)}$$

We have:

$$-(\mathbf{A}^*\Sigma_p^{-1}\mathbf{A})^{-1}\frac{\partial}{\partial \mathbf{m}}(\mathbf{A}^*\Sigma_p^{-1}\mathbf{A})(\mathbf{A}^*\Sigma_p^{-1}\mathbf{A})^{-1}=-\mathbf{A}^{-1}\Sigma_p\mathbf{A}^{-*}\mathbf{A}^*\Sigma_p^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{A}^{-1}\Sigma_p\mathbf{A}^{-*}, \quad \text{(C-8)}$$

$$-\mathbf{A}^{-1}\Sigma_p\mathbf{A}^{-*}\mathbf{A}^*\Sigma_p^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{A}^{-1}\Sigma_p\mathbf{A}^{-*}=-\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{A}^{-1}\Sigma_p\mathbf{A}^{-*}. \quad \text{(C-9)}$$

Finally, we obtain:

$$\frac{\partial \Sigma(\mathbf{m})}{\partial \mathbf{m}}=-\mathbf{P}_i\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{A}^{-1}\Sigma_p\mathbf{A}^{-*}\mathbf{P}_i^*=-(\mathbf{P}_i\mathbf{A}^{-1})\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{A}^{-1}\Sigma_p(P_i\mathbf{A}^{-1})^*. \quad \text{(C-10)}$$

REFERENCES

Abubakar, A., Hu, W., van den Berg, P. M., & Habashy, T. M., 2008. A finite-difference contrast source inversion method, Inverse problems, 24(6), 1–17.

Aghamiry, H. S., A. Gholami, and S. Operto, 2018, Improving full waveform inversion by wavefield reconstruction with the alternating direction method of multipliers: Geophysics, 1–79.

———, 2019, Accurate and efficient data-assimilated wavefield reconstruction in the time domain: Geophysics, **82**, A7.

———, 2020, Attenuation imaging by wavefield reconstruction inversion with bound constraints and total variation regularization: Geophysics, **85**, 1–54.

Beydoun, W. B., and A. Tarantola, 1988, First born and rytov approximations: Modeling and inversion conditions in a canonical example: The Journal of the Acoustical Society of America, **83**, 1045–1055.

Bezdek, J. C., and R. J. Hathaway, 2002, Some notes on alternating optimization: Presented at the AFSS International Conference on Fuzzy Systems.

Billette, F. J., and Brandsberg-Dahl, S, 2005, The 2004 BP velocity benchmark: 67th European Association of Geoscientists & Engineers Annual International Meeting, EAGE, Expanded Abstracts.

Bozda˘g, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: Geophysical Journal International, **185**, 845–870.

Brittan, J., and I. Jones, 2019, Fwi evolution—from a monolith to a toolkit: The Leading Edge, **38**, 179–184.

Brossier, R., S. Operto, and J. Virieux, 2009, Seismic imaging of complex onshore structures by 2d elastic frequency-domain full-waveform inversion: Geophysics, **74**, WCC105.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Choi, Y., and T. Alkhalifah, 2015, Unwrapped phase inversion with an exponential damping: Geophysics, **80**, 251–264.

da Silva, N. V., 2015, Automatic Wavefield Reconstruction Inversion: 85th Annual International Meeting, SEG, Expanded Abstracts, 1332–1336.

Fu, L., and W. W. Symes, 2017a, An adaptive multiscale algorithm for efficient extended waveform inversion: Geophysics, **82**, R183–R197.

———, 2017b, A discrepancy-based penalty method for extended waveform inversion: Geophysics, **82**, R287–R298.

Fabien-Ouellet, G., Gloaguen, E., and Giroux, B, 2017, A stochastic L-BFGS approach for full-waveform inversion: 87th SEG Technical Program Expanded Abstracts, 1622-1627.

Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: numerical results: Geophysics, 51, 1387-1403.

Gee, L. S., and T. H. Jordan, 1992, Generalized seismological data functionals: Geophysical Journal International, **111**, 363–390.

Gholami, A., H. S. Aghamiry, and S. Operto, 2021a, Clarifying some issues on extended fwi: scattered-field equation, time reversal and source reconstruction: SEG/AAPG/SEPM First International Meeting for Applied Geoscience & Energy.

———, 2021b, Extended full waveform inversion in the time domain by the augmented lagrangian method: ArXiv, abs/2011.14102.

Guasch, L., M. Warner, and C. Ravaut, 2019, Adaptive waveform inversion: Practice: Geophysics, **84**, R447–R461.

Golub, G., and Pereyra, V., 2003, Separable nonlinear least squares: the variable projection method and its applications. Inverse problems, **19**(2), R1.

Ha, T., W. Chung, and C. Shin, 2009, Waveform inversion using a back-propagation algorithm and a huber function norm: Geophysics, **74**, R15–R24.

Huang, G., R. Nammour, and W. Symes, 2017, Full-waveform inversion via source-receiver extension: Geophysics, **82**, R153–R171.

Huang, G., R. Nammour, and W. W. Symes, 2018, Source-independent extended waveform inversion based on space-time source extension: Frequency-domain implementation: Geophysics, **83**, R449–R461.

Li, Z.C., Y.Z. Lin, K. Zhang, Y.Y. Li, and Z.N. Yu, 2017, Time-domain wavefield reconstruction inversion: Applied Geophysics, **14**, 523–528.

Lin, Y., Z. Li, and K. Zhang, 2018, Unwrapped instantaneous-phase inversion based on new decompose method: 88th Annual International Meeting, SEG, Expanded Abstracts, 1369–1373.

Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653.

Louboutin, M., G. Rizzuti, and F. J. Herrmann, 2020, Time-domain wavefield reconstruction inversion in a TTI medium: arXiv preprint arXiv:2004.07355.

Mulder, W., 2013, Subsurface offset behavior in velocity analysis with extended reflectivity images: Geophysical Prospecting, **62**, 17–33.

M´etivier, L., R. Brossier, Q. M´erigot, E. Oudet, and J. Virieux, 2017, Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion: Geophysical Journal International, **205**, 332–364.

Plessix, R.-E., G. Baeten, J. W. de Maag, M. Klaassen, Z. Rujie, and T. Zhifei, 2010, Application of acoustic full waveform inversion to a low-frequency large-offset land data set: 80th Annual International Meeting, SEG, Expanded Abstracts, 930–934.

Pratt, R., L. Sirgue, B. Hornby, and J. Wolfe, 2008, Cross-well waveform tomography in fine-layered sediments—meeting the challenges of anisotropy: 70th European Association of Geoscientists & Engineers Annual International Meeting, EAGE, Expanded Abstracts, F020.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888–901.

Qiu, L., J. Ramos-Mart´ınez, A. Valenciano, J. Kirkebø, and N. Chemingui, 2017, Mitigating the cycle-skipping of full-waveform inversion: An optimal transport approach with exponential encoding: Seg 2017 Workshop: Full-Waveform Inversion and Beyond, Beijing, China, 20-22 November, 1–4.

Rizzuti, G., M. Louboutin, R. Wang, E. Daskalakis, and F. Herrmann, 2019, A dual formulation for time-domain wavefield reconstruction inversion: 89th SEG Technical Program Expanded Abstracts,1480–1485.

Rizzuti, G., M. Louboutin, R. Wang, and F. Herrmann, 2021, A dual formulation of wavefield reconstruction inversion for large-scale seismic inversion: Geophysics, **86**(6), R879-893.

Shin, C., and Y. Ho Cha, 2009, Waveform inversion in the Laplace—Fourier domain: Geophysical Journal International, **177**, 1067–1079.

Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Symes, W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical prospecting, **56**, 765–790.

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

———, 2005, Inverse problem theory and methods for model parameter estimation: siam, 89.

van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667.

———, 2015, A penalty method for PDE-constrained optimization in inverse problems: Inverse Problems, **32**, 015007.

Van Leeuwen, T., and W. Mulder, 2010, A correlation-based misfit criterion for wave-equation traveltime tomography: Geophysical Journal International, **182**, 1383–1394.

van Leeuwen, T., 2019, A note on extended full waveform inversion: arXiv preprint, arXiv:1904.00363.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Wang, C., D. Yingst, P. Farmer, and J. Leveille, 2016, Full-waveform inversion with the reconstructed wavefield method: 86th SEG Technical Program Expanded Abstracts, 1237–1241.

Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: Geophysics, **81**, R429–R445.

Weglein, A. B., 2013, A timely and necessary antidote to indirect methods and so-called p-wave fwi: The Leading Edge, **32**, 1192–1204.

Yong, P., J. Huang, Z. Li, and W. Liao, 2018, Earth mover's distance based on gradient flow for full waveform inversion: 80th European Association of Geoscientists & Engineers Annual International Meeting, EAGE, Expanded Abstracts, 1-5.

Yuzhao. Lin, Huaishan. Liu, Lei. Xing and Haoran. Lin, 2022, Time-Domain Wavefield Reconstruction Inversion Solutions in the Weighted Full Waveform Inversion Form: IEEE Transactions on Geoscience and Remote Sensing, **60**, 1-14.

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS