

Time-domain elastic Gauss–Newton full-waveform inversion: a matrix-free approach

Ke Chen¹ and Mauricio D. Sacchi

Department of Physics, University of Alberta, Edmonton, AB T6G 2E1, Canada. E-mail: ke7@ualberta.ca

Accepted 2020 June 29. Received 2020 May 17; in original form 2020 January 3

SUMMARY

We present a time-domain matrix-free elastic Gauss–Newton full-waveform inversion (FWI) algorithm. Our algorithm consists of a Gauss–Newton update with a search direction calculated via elastic least-squares reverse time migration (LSRTM). The conjugate gradient least-squares (CGLS) method solves the LSRTM problem with forward and adjoint operators derived via the elastic Born approximation. The Hessian of the Gauss–Newton method is never explicitly formed or saved in memory. In other words, the CGLS algorithm solves for the Gauss–Newton direction via the application of implicit-form forward and adjoint operators which are equivalent to elastic Born modelling and elastic reverse time migration, respectively. We provide numerical examples to test the proposed algorithm where we invert for P - and S -wave velocities simultaneously. The proposed algorithm performs positively on mid-size problems where we report solutions of slight improvement than those computed using the conventional non-linear conjugate gradient method. In spite of the aforementioned limited gain, the theory developed in this paper contributes to a better understanding of time-domain elastic Gauss–Newton FWI.

Key words: Inverse theory; Waveform inversion; Computational seismology; Seismic tomography; Wave propagation.

1 INTRODUCTION

Full-waveform inversion (FWI) consists of quantitatively estimating subsurface parameters using observed seismic waveform data. FWI requires solving an optimization problem where a non-linear objective function is minimized. The cost function measures the difference between observed and numerically modelled seismic data. The development of FWI starts in the early 1980s with the original work of Bamberger *et al.* (1979 and 1982) where the adjoint-state method of optimal control theory (Lions 1971) is applied to invert for the coefficients of the 1-D acoustic wave equation. On a similar vein, Tarantola (1984) formulates time-domain acoustic FWI as a non-linear least-squares inverse problem which is solved using a gradient-based iterative algorithm (Tarantola & Valette 1982). Numerical examples for 2-D time-domain acoustic FWI are provided by Gauthier *et al.* (1986). To better approximate the realistic earth, elasticity is further incorporated in FWI by several authors (Tarantola 1986; Mora 1987a, 1988; Sears *et al.* 2008). Moreover, the development of high-performance computing in recent years has enabled 3-D time-domain elastic FWI at exploration scale (Guasch *et al.* 2012; Vigh *et al.* 2014; Borisov & Singh 2015; Borisov *et al.* 2018; Oh *et al.* 2018; Trinh *et al.* 2018) and global scale (Fichtner *et al.* 2009; Tape *et al.* 2009, 2010). In the last decade, geophysical prospecting of oil and gas has adopted FWI for estimating velocity models (Xu *et al.* 2012; Brossier *et al.* 2015; Luo *et al.* 2016; Mao *et al.* 2016; Wang *et al.* 2016; Brandsberg-Dahl *et al.* 2017), high-resolution imaging (Plessix & Perkins 2010; Sirgue *et al.* 2010; Baumstein 2014; Lu 2016; Shen *et al.* 2018) and time-lapse reservoir monitoring (Routh *et al.* 2012; Raknes & Arntsen 2014; Alemie & Sacchi 2016; Yang *et al.* 2016; Kamei & Lumley 2017). Conventional time-domain FWI algorithms are generally implemented via gradient-based methods, such as the method of steepest descent, the non-linear conjugate gradient method or quasi-Newton methods. The Gauss–Newton method and the Newton’s method normally converge faster than gradient-based techniques, a characteristic reported in numerous studies (Pratt *et al.* 1998; Santosa & Symes 1988; Gratton *et al.* 2007; Burstedde & Ghattas 2009).

Gauss–Newton and Newton’s methods have been around for a long time for solving non-linear least-squares inverse problems. For instance, Santosa & Symes (1988) develop a truncated Newton method for the 1-D time-domain acoustic FWI problem. In each Newton’s iteration, the Hessian can be implicitly inverted using the conjugate gradient method. The Gauss–Newton method is a simplified case of Newton’s method where the second derivative is discarded (Santosa & Symes 1988). Akcelik *et al.* (2002) implement a 3-D time-domain acoustic Gauss–Newton–Krylov FWI. The Gauss–Newton Hessian is implicitly inverted using the symmetric conjugate gradient algorithm. Epanomeritakis *et al.* (2008) extend the Gauss–Newton–Krylov FWI to the 3-D time-domain elastic case, but only a single parameter

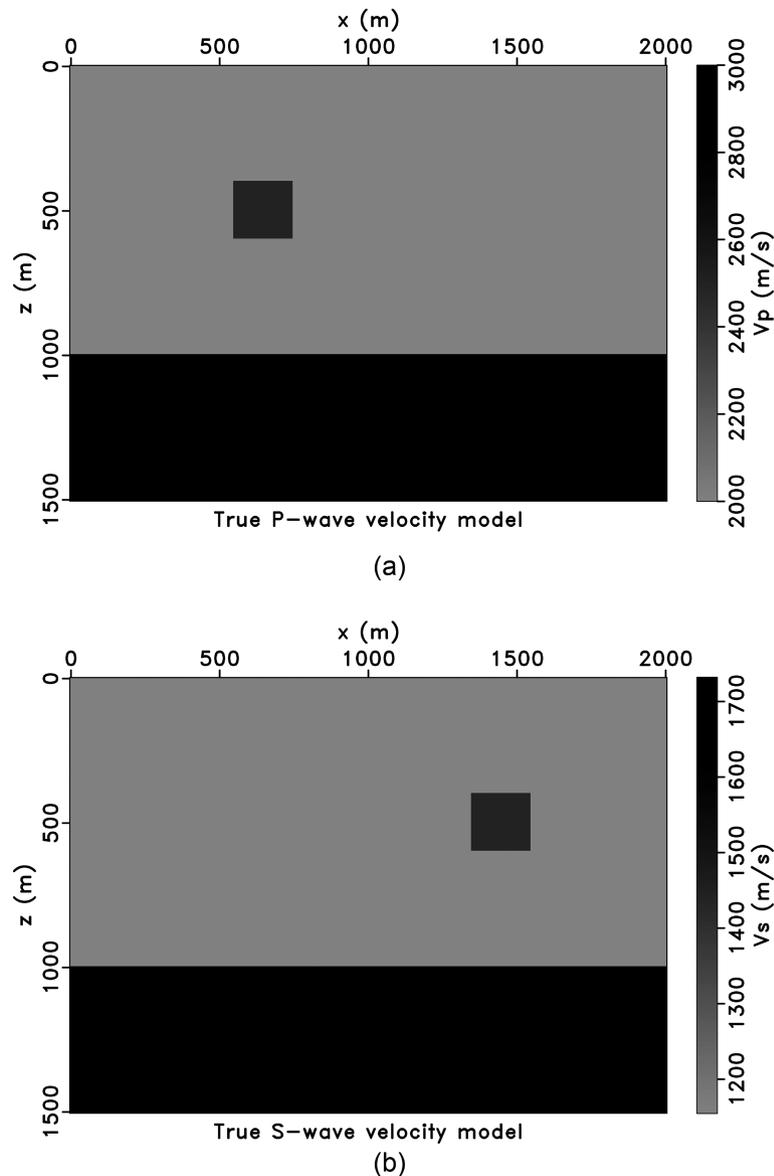


Figure 1. Elastic inclusion model. (a) True P -wave velocity model. (b) True S -wave velocity model.

(shear modulus) is inverted in their numerical examples. Burstedde & Ghattas (2009) investigate a 1-D time-domain acoustic FWI using the truncated Newton method proposed in Santosa & Symes (1988) and recommend early termination of the conjugate gradient to avoid directions of negative curvature. Recently, Yang *et al.* (2018) present a 2-D time-domain truncated Newton FWI for visco-acoustic media. In the frequency-space domain, Metivier *et al.* (2013, 2014, 2017) and Anagaw & Sacchi (2018a) apply the truncated Newton method to the 2-D acoustic FWI problem. Li *et al.* (2016) also investigate an ℓ_1 norm regularized Gauss–Newton FWI method in frequency-space domain. More recently, Witte *et al.* (2018) review different optimization methods for FWI in frequency-space domain.

It is worth to mention a more straightforward way to implement Gauss–Newton and Newton’s methods by explicitly computing, storing and inverting the Hessian matrix (Pratt *et al.* 1998; Sheen *et al.* 2006; Chen *et al.* 2007; Pan *et al.* 2016). However, this approach is too computationally expensive and unfeasible for application to large-scale problems.

The direct solver for the wave equation in the frequency domain requires large computer memory. This becomes the principal bottleneck for applying frequency-domain FWI methods in 3-D large scale problems (Vigh & Starr 2008b). Also, the frequency-domain methods do not allow flexible time axis related operations. Hence, the time-domain techniques are becoming more and more popular in the field on FWI. The material present in our paper is in time-space domain.

In this paper, we present a time-domain matrix-free elastic Gauss–Newton FWI algorithm (Chen & Sacchi 2018). Compared to Newton’s method, the Gauss–Newton method for solving the non-linear least-squares inverse problem does not require the computation of the second derivatives of the misfit function. One of the advantages of the Gauss–Newton method over Newton’s method is that its Hessian is always positive definite which is necessary for local convergence (Gratton *et al.* 2007). Our matrix-free Gauss–Newton algorithm is a nested process: the outer loop is the Gauss–Newton update (Nocedal & Wright 2006) and the inner loop is the conjugate gradient least-squares (CGLS)

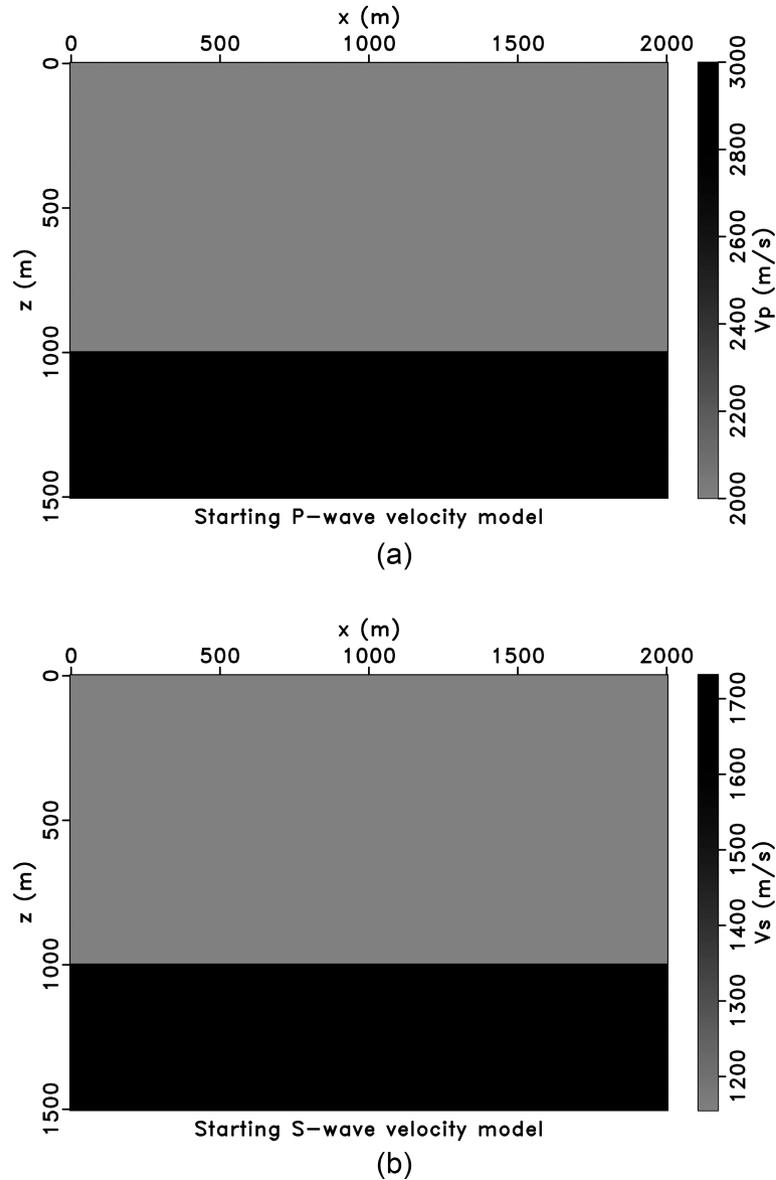


Figure 2. (a) Starting P -wave velocity model. (b) Starting S -wave velocity model.

Table 1. Parameters used in the elastic inclusion model inversion example.

Frequency group	Stopping criterion for FWI loop ϵ	Stopping criterion for CGLS loop ξ
3–5 Hz	0.01	0.5
3–10 Hz	0.05	0.5
3–15 Hz	0.2	0.5
3–30 Hz	0.2	0.5

method (Hestenes & Stiefel 1952; Paige & Saunders 1982) that estimates the Gauss–Newton search direction. We recognize that the inner loop is equivalent to applying an elastic least-squares reverse time migration (LSRTM) to data residuals (Chen & Sacchi 2017; Feng & Schuster 2017; Duan *et al.* 2017; Ren *et al.* 2017). The Fréchet derivative operator acts as the elastic Born modelling operator and the adjoint of Fréchet derivative operator acts as the elastic RTM operator. In the CGLS algorithm, the step size is analytically calculated without of the need for a line search. The CGLS method is also preconditioned using an elastic pseudo-Hessian operator (Shin *et al.* 2001). The presented algorithm is matrix-free; it only requires the forward Fréchet derivative, and the adjoint Fréchet derivative operators applied in implicit forms to vectors. The operators are applied on vectors efficiently via the adjoint-state method (Lions 1971; Tromp *et al.* 2005; Fichtner *et al.* 2006; Plessix 2006). Once we obtain the Gauss–Newton search direction, the algorithm uses a line search based on parabolic fitting (Vigh *et al.* 2009) to estimate the step length. The proposed algorithm is adopted to invert for P - and S -wave velocities simultaneously. We develop the whole theory in 3-D, but we present numerical examples in 2-D which conform to our current computational resources.

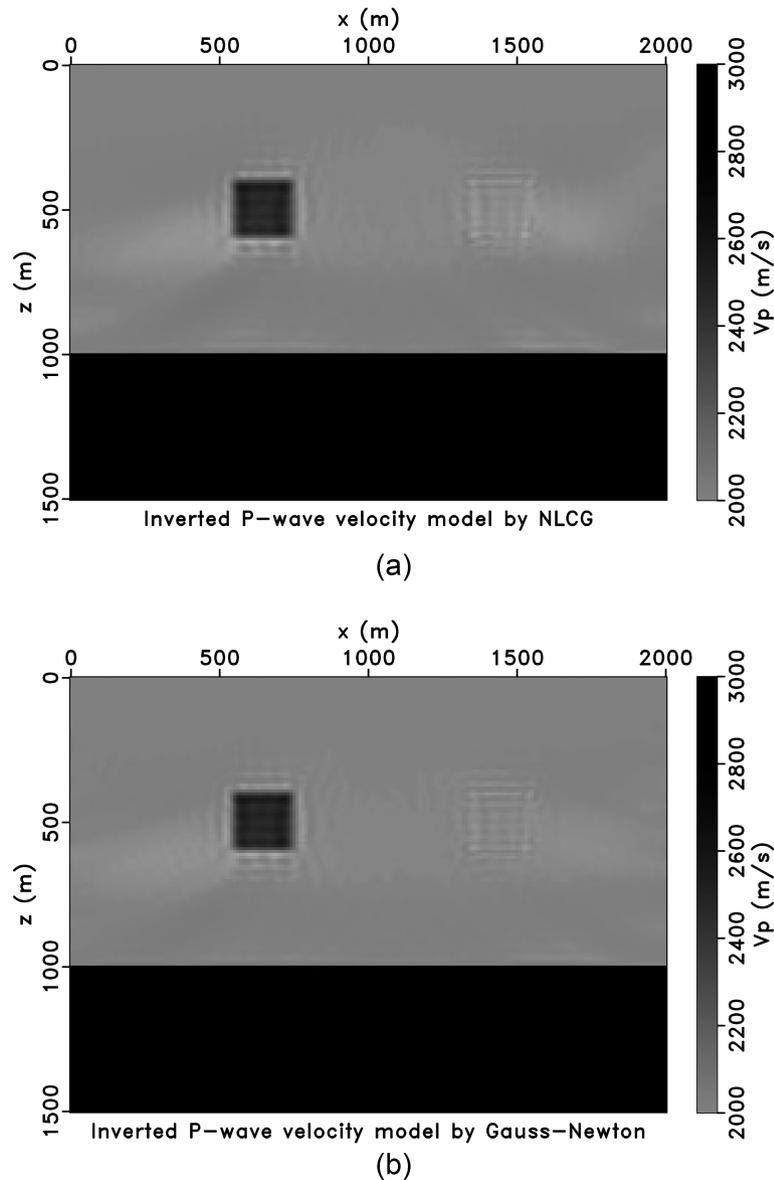


Figure 3. (a) Inverted P -wave velocity model by elastic NLCG FWI. (b) Inverted P -wave velocity model by elastic Gauss–Newton FWI.

Our research complements previous work on the application of the Gauss–Newton method to FWI problem. It also provides numerical analysis tests for a better understanding of the algorithm. We stress the following differences between our work and the work mentioned above and summarize our contributions in the next paragraph. First, our research is focused on elastic parameters inversion using the truncated Gauss–Newton method. For instance, Gauss–Newton optimization is also adopted by Yang *et al.* (2018) and Metivier *et al.* (2013). However, we stress that Yang *et al.* (2018)’s analysis pertains to waveform inversion in visco-acoustic media and Metivier *et al.* (2013) investigate single parameter inversion in acoustic media via frequency-domain FWI. Closer to our work, Epanomeritakis *et al.* (2008) present a Newton-CG method for elastic FWI. However, Epanomeritakis *et al.* (2008) only invert for one parameter in their numerical examples, the shear modulus. Moreover, their inversion is for a lithospheric scale problem. To the best of our knowledge, our paper is the first presentation of the numerical results of simultaneous inversion of P - and S -wave velocities using the matrix-free truncated Gauss–Newton method in the time domain. Other contributions worth mentioning include the following. We estimate the Gauss–Newton direction using the CGLS method (Paige & Saunders 1982) and explicitly point out the connection between elastic LSRTM and the elastic Gauss–Newton FWI. This framework offers considerable freedom to design and apply Gauss–Newton FWI algorithms from existing LSRTM codes that nowadays are becoming popular in the exploration seismology community (Anikiev *et al.* 2013; Xu *et al.* 2016; Chen & Sacchi 2017; Duan *et al.* 2017; Feng & Schuster 2017; Gu *et al.* 2017; Ren *et al.* 2017; Sun *et al.* 2018). We also point out that we adopt the optimize-then-discretize approach (Borzi & Schulz 2012) for designing of the proposed algorithm. The derivation of the adjoint-state equation, the Fréchet derivative and its adjoint are in continuous functional space form and the resulted formulations are discretized later to develop a numerical code.

Our work has lead to several interesting observations. The proposed elastic Gauss–Newton FWI generates slightly better-inverted models than the elastic non-linear conjugate gradient FWI for inverting localized velocity anomalies. The computational cost of the proposed method

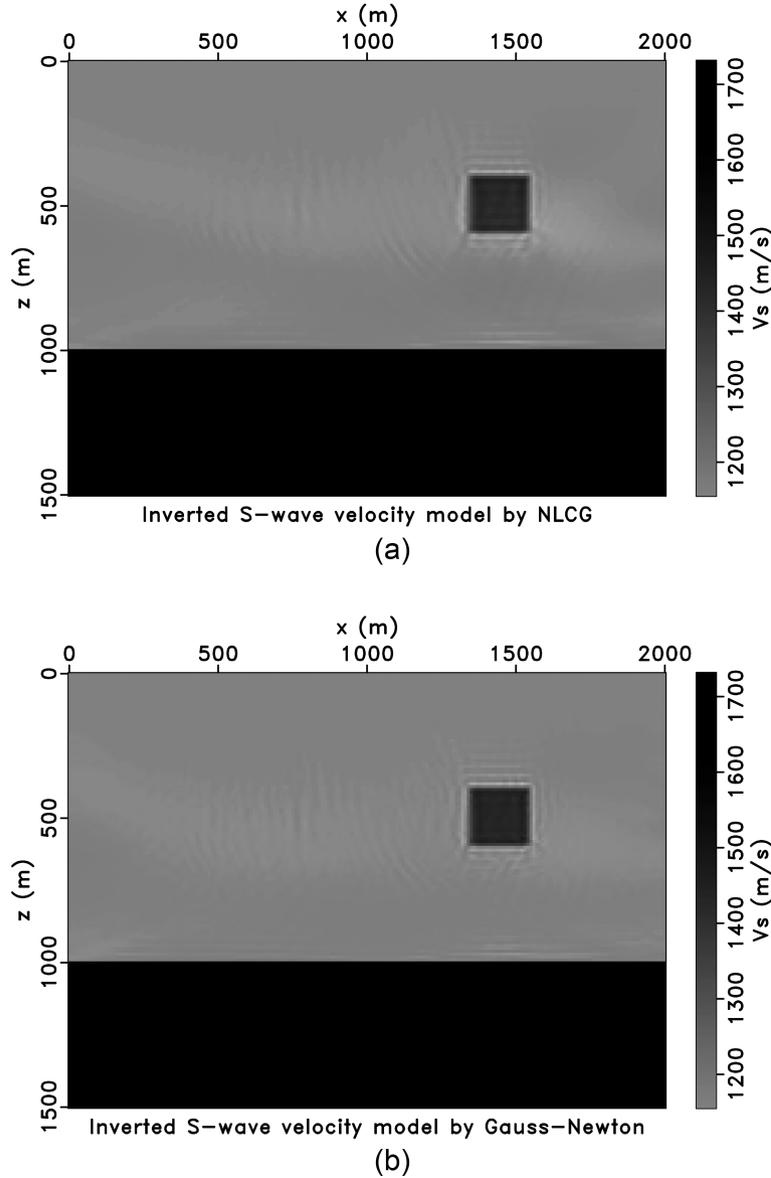


Figure 4. (a) Inverted S -wave velocity model by elastic NLCG FWI. (b) Inverted S -wave velocity model by elastic Gauss–Newton FWI.

is tolerable for mid-scale problems. We recognize that the proposed technique does not provide significant improvement when compared to the non-linear conjugate gradient method for simultaneously inverting for P- and S-wave velocities. The latter is particularly true when comparing smooth background model parameters estimated by our Gauss–Newton algorithm and the non-linear conjugate gradient method. For large-scale computational problems, we believe that the proposed Gauss–Newton solver might not be the best alternative for time-domain elastic FWI because similar results could be obtained via a less expensive solver such as non-linear conjugate gradient. However, we understand that last comment is debatable and ongoing research with more exhaustive tests is needed to fully address the advantage of adopting Gauss–Newton solvers for time-domain elastic FWI. For example, the elastic FWI could begin with non-linear conjugate gradient algorithm and switch over to Gauss–Newton method in the process considering the trade-off between convergence property gain and computational cost of the Gauss–Newton algorithm (Burstedde & Ghattas 2009). Moreover, the crosstalk between density and velocities is more severe than the crosstalk between P- and S-velocities in the elastic FWI. The Gauss–Newton algorithm has the potential to reduce the crosstalk between the density and velocities in elastic FWI which provides a future research direction. From a more general perspective, our study nevertheless formalizes connections to current work in the field of elastic migration via LSRTM.

This paper is organized as follows: First, we describe the system of equations that we have adopted to forward model elastic wavefields. We then discuss the general time-domain elastic Gauss–Newton FWI algorithm. Subsequently, we formulate the Gauss–Newton search direction as the solution of an elastic LSRTM problem. Then, we solve the Gauss–Newton search direction using CGLS with preconditioning and summarize the time-domain matrix-free elastic Gauss–Newton FWI algorithm. In the last section, we provide numerical examples to evaluate the proposed algorithm.

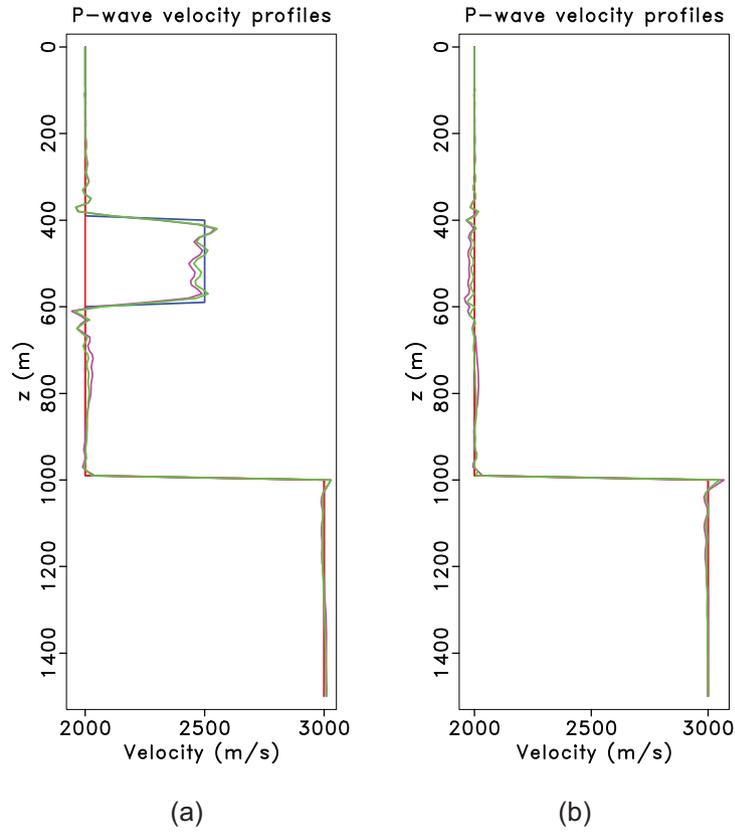


Figure 5. Profiles of inverted *P*-wave velocity models at (a) $x = 650$ m and (b) $x = 1450$ m. Blue: true model; Red: starting model; Purple: inverted model by elastic NLCG FWI; Green: inverted model by elastic Gauss-Newton FWI.

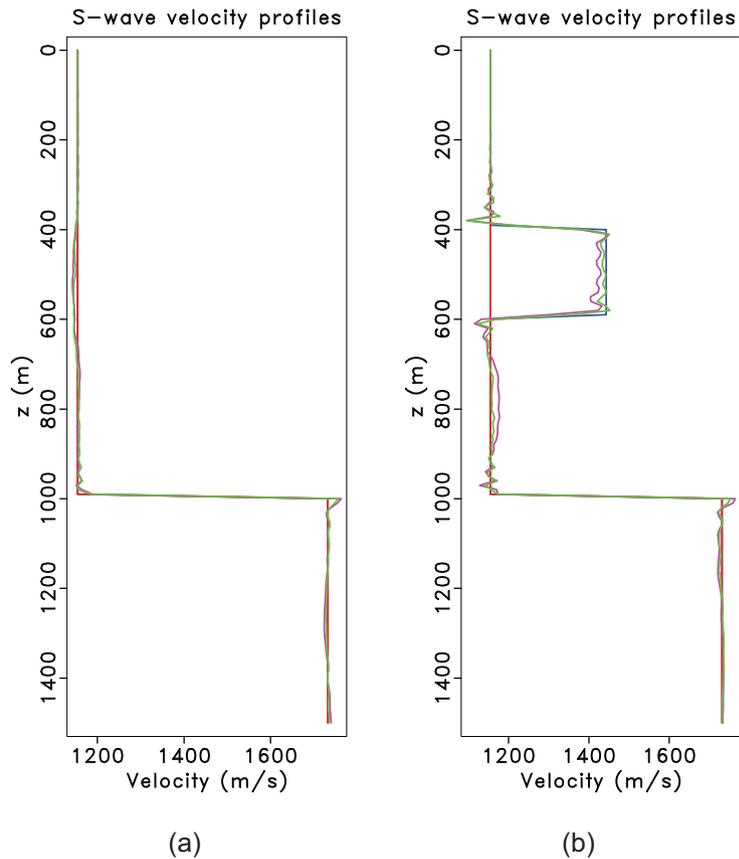


Figure 6. Profiles of inverted *S*-wave velocity models at (a) $x = 650$ m and (b) $x = 1450$ m. Blue: true model; Red: starting model; Purple: inverted model by elastic NLCG FWI; Green: inverted model by elastic Gauss-Newton FWI.

2 THEORY

2.1 The forward problem in the time-space domain

The propagation of seismic waves in heterogeneous, isotropic, elastic earth media can be expressed by the elastodynamic equations (Virieux 1986; Graves 1996)

$$\left\{ \begin{aligned} \rho \frac{\partial v_x}{\partial t} - \left(\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + \frac{\partial \sigma_{xz}}{\partial z} \right) &= 0, \\ \rho \frac{\partial v_y}{\partial t} - \left(\frac{\partial \sigma_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + \frac{\partial \sigma_{yz}}{\partial z} \right) &= 0, \\ \rho \frac{\partial v_z}{\partial t} - \left(\frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{yz}}{\partial y} + \frac{\partial \sigma_{zz}}{\partial z} \right) &= 0, \\ \frac{\partial \sigma_{xx}}{\partial t} - (\lambda + 2\mu) \frac{\partial v_x}{\partial x} - \lambda \left(\frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right) &= f_{\sigma_{xx}}, \\ \frac{\partial \sigma_{yy}}{\partial t} - (\lambda + 2\mu) \frac{\partial v_y}{\partial y} - \lambda \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_z}{\partial z} \right) &= f_{\sigma_{yy}}, \\ \frac{\partial \sigma_{zz}}{\partial t} - (\lambda + 2\mu) \frac{\partial v_z}{\partial z} - \lambda \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} \right) &= f_{\sigma_{zz}}, \\ \frac{\partial \sigma_{yz}}{\partial t} - \mu \left(\frac{\partial v_y}{\partial z} + \frac{\partial v_z}{\partial y} \right) &= f_{\sigma_{yz}}, \\ \frac{\partial \sigma_{xz}}{\partial t} - \mu \left(\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x} \right) &= f_{\sigma_{xz}}, \\ \frac{\partial \sigma_{xy}}{\partial t} - \mu \left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right) &= f_{\sigma_{xy}}, \end{aligned} \right. \quad (1)$$

where (v_x, v_y, v_z) are the horizontal and vertical particle velocity fields, $(\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{yz}, \sigma_{xz}, \sigma_{xy})$ are the stress fields, $(f_{\sigma_{xx}}, f_{\sigma_{yy}}, f_{\sigma_{zz}}, f_{\sigma_{yz}}, f_{\sigma_{xz}}, f_{\sigma_{xy}})$ are the source terms, ρ is density, λ and μ are the Lamé parameters. These coefficients describe the spatially variable property of the earth. The latter is related to the seismic P - and S -wave velocities via $\lambda + 2\mu = \rho V_p^2$ and $\mu = \rho V_s^2$. To make the notation concise, we drop the dependency on spatial and temporal coordinates \mathbf{x} and t of our variables [i.e. $\rho = \rho(\mathbf{x})$ and $v_x = v_x(\mathbf{x}, t)$]. The elastic wave equation complies the zero initial condition, the free surface boundary condition on the surface of the modelled earth ($z = 0$), and absorbing boundary conditions on the sides and bottom of the model. To make our analysis more clear, the elastic wave equation given by eq. (1) can be organized as (Vigh *et al.* 2014)

$$\left\{ \begin{aligned} \rho \frac{\partial \mathbf{v}}{\partial t} - \mathbf{D}\boldsymbol{\sigma} &= \mathbf{0}, \\ \frac{\partial \boldsymbol{\sigma}}{\partial t} - \mathbf{C}\mathbf{D}^T \mathbf{v} &= \mathbf{f}_\sigma, \end{aligned} \right. \quad (2)$$

where

$$\mathbf{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}, \quad \boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{pmatrix}, \quad \mathbf{f}_\sigma = \begin{pmatrix} f_{\sigma_{xx}} \\ f_{\sigma_{yy}} \\ f_{\sigma_{zz}} \\ f_{\sigma_{yz}} \\ f_{\sigma_{xz}} \\ f_{\sigma_{xy}} \end{pmatrix}, \quad (3)$$

$$\mathbf{C} = \begin{pmatrix} \lambda + 2\mu & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda + 2\mu & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & \lambda + 2\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} \frac{\partial}{\partial x} & 0 & 0 & 0 & \frac{\partial}{\partial z} & \frac{\partial}{\partial y} \\ 0 & \frac{\partial}{\partial y} & 0 & \frac{\partial}{\partial z} & 0 & \frac{\partial}{\partial x} \\ 0 & 0 & \frac{\partial}{\partial z} & \frac{\partial}{\partial y} & \frac{\partial}{\partial x} & 0 \end{pmatrix},$$

where \mathbf{v} denotes the particle velocity field, $\boldsymbol{\sigma}$ is the stress field, \mathbf{C} is the isotropic elastic tensor in Voigt notation, \mathbf{D} is a collection of spatial differential operators, and \mathbf{f}_σ is the source term. The partial differential eqs (2) can be expressed as

$$\begin{pmatrix} \rho \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \frac{\partial}{\partial t} \mathbf{u} - \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C}\mathbf{D}^T & \mathbf{0} \end{pmatrix} \mathbf{u} = \mathbf{f}, \quad (4)$$

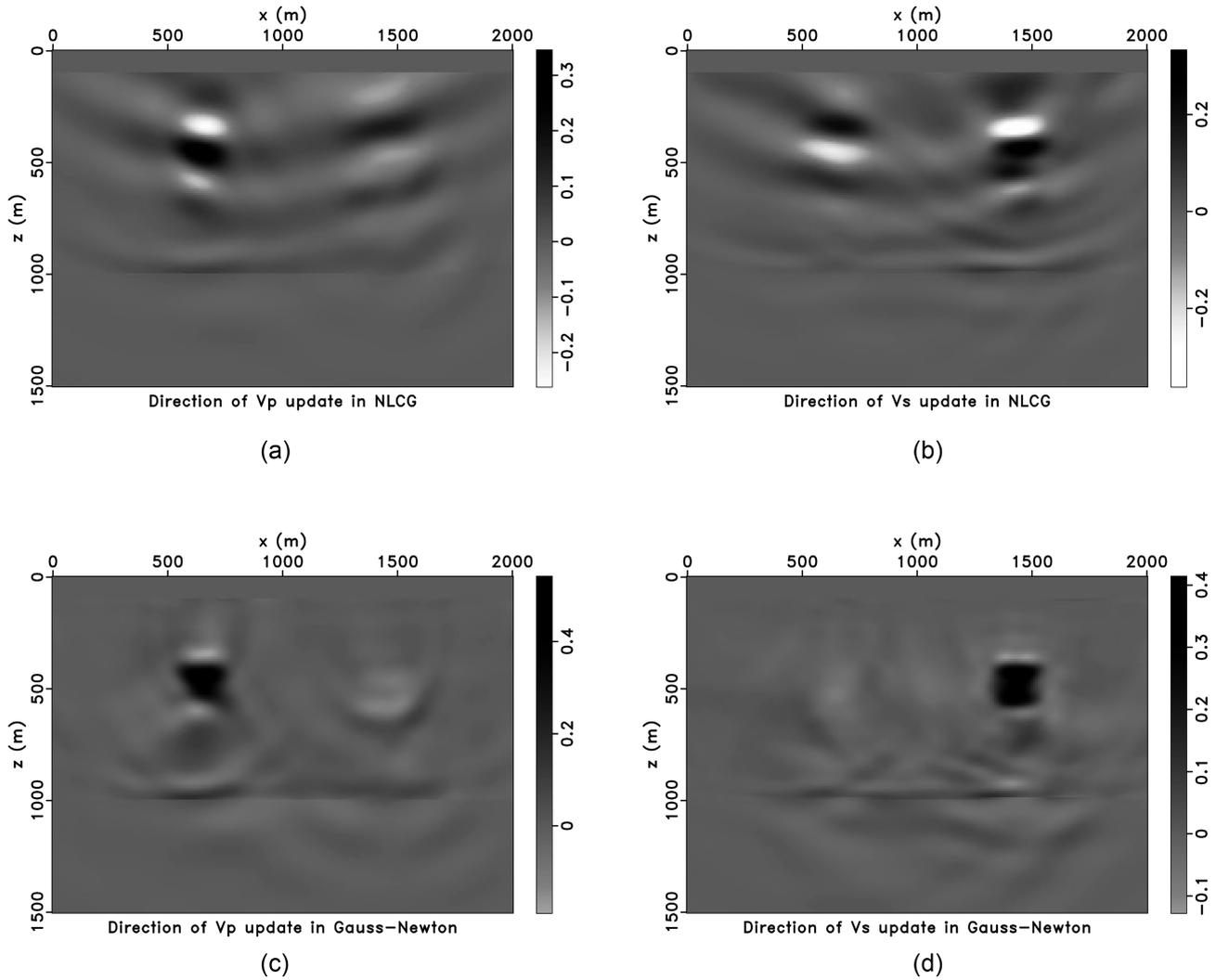


Figure 7. The model update direction in the first non-linear FWI iteration in the first frequency band (3–5 Hz). (a) Direction of V_p update in elastic NLCG FWI (negative of gradient direction). (b) Direction of V_s update in elastic NLCG FWI. (c) Direction of V_p update in elastic Gauss–Newton FWI (solved by CGLS). (d) Direction of V_s update in elastic Gauss–Newton FWI. Figures are plotted with original scale.

where $\mathbf{u} = \begin{pmatrix} \mathbf{v} \\ \sigma \end{pmatrix}$ is the wavefield, $\mathbf{f} = \begin{pmatrix} \mathbf{0} \\ \mathbf{f}_\sigma \end{pmatrix}$ is the source and \mathbf{I} is identity matrix. Note that, the wavefield \mathbf{u} in eq. (4) depends on the spatial and temporal coordinates \mathbf{x} and t , that is $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$.

A more general representation of the wave equation can be written as

$$\mathbf{S}\mathbf{u} = \mathbf{f}, \quad (5)$$

where \mathbf{u} is the wavefield vector in space \mathbb{U} , \mathbf{f} is the source vector in space \mathbb{F} , \mathbf{S} is the wave equation operator mapping \mathbb{U} to \mathbb{F} . The wave equation operator is a function of media parameters. Hence, we denote $\mathbf{S} = \mathbf{S}(\mathbf{m})$, where \mathbf{m} represents model parameters. The model vector \mathbf{m} belongs to the model space \mathbb{M} . The solution of the wave equation can be abstractly expressed as

$$\mathbf{u} = \mathbf{S}^{-1}\mathbf{f} = \mathbf{G}\mathbf{f}, \quad (6)$$

where $\mathbf{G} = \mathbf{S}^{-1}$ is the inverse of the wave equation operator \mathbf{S} and is called the Green's operator (Tarantola 1988). The wavefield \mathbf{u} is a non-linear implicit function of the vector of model parameters \mathbf{m}

$$\mathbf{u} = \mathbf{u}(\mathbf{m}). \quad (7)$$

Seismic data \mathbf{d} are time series recorded by receivers deployed on the surface of the earth and the recording process entails extracting seismograms from \mathbf{u} . This operator can be represented via the following expression

$$\mathbf{d} = \mathbf{R}\mathbf{u}, \quad (8)$$

where \mathbf{R} is the sampling operator also called the restriction operator.

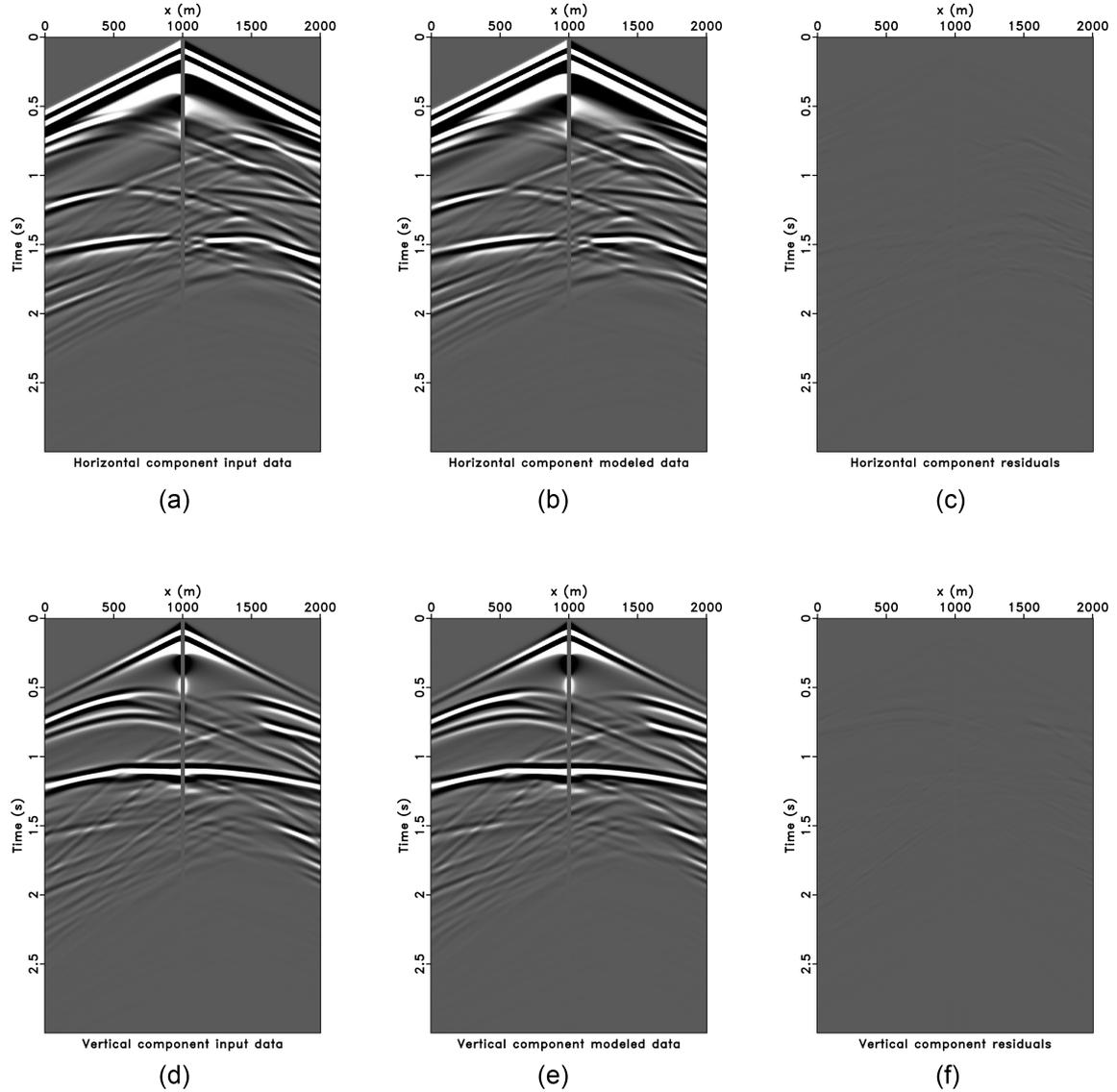


Figure 8. Data and residuals for elastic NLCG FWI. (a) Observed horizontal component data of shot at $x = 1000$ m. (b) Horizontal component data modelled by elastic NLCG FWI inverted models. (c) Horizontal component data residual. (d) Observed vertical component data of shot at $x = 1000$ m. (e) Vertical component data modelled by elastic NLCG FWI inverted models. (f) Vertical component data residual.

2.2 General formulation of time-domain elastic Gauss–Newton FWI

Elastic FWI estimates the distribution of subsurface elastic parameters from the observed seismic data. It is usually formulated as a non-linear least-squares inverse problem that minimizes the functional (Tarantola 1984, 1986)

$$J(\mathbf{m}) = \frac{1}{2} \sum_{i=1}^{N_s} \|\mathbf{d}_i(\mathbf{m}) - \mathbf{d}_i^{obs}\|_2^2, \quad (9)$$

where $\mathbf{d}_i(\mathbf{m})$ represents the forward modelled seismic data for the i th seismic source. The vector \mathbf{d}_i^{obs} denotes the observed seismic data for the i th source, and $\|\cdot\|_2$ denotes the ℓ_2 norm of a vector. The non-linear inverse problem belongs to a class often called the partial differential equations constrained optimization problem (Lions 1971; Borzi & Schulz 2012). Given the large size of the model space, global optimization methods are too expensive for solving FWI problems. Hence, FWI is usually solved via iterative local optimization algorithms.

Introducing a model perturbation $\mathbf{m} \rightarrow \mathbf{m} + \delta\mathbf{m}$ and a second-order Taylor series expansion, the objective function in the vicinity of \mathbf{m} is given by the following expression (Virieux & Operto 2009)

$$J(\mathbf{m} + \delta\mathbf{m}) = J(\mathbf{m}) + \delta\mathbf{m}^T \frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} + \frac{1}{2} \delta\mathbf{m}^T \frac{\partial^2 J(\mathbf{m})}{\partial \mathbf{m}^2} \delta\mathbf{m} + \mathcal{O}(\|\delta\mathbf{m}\|^3), \quad (10)$$

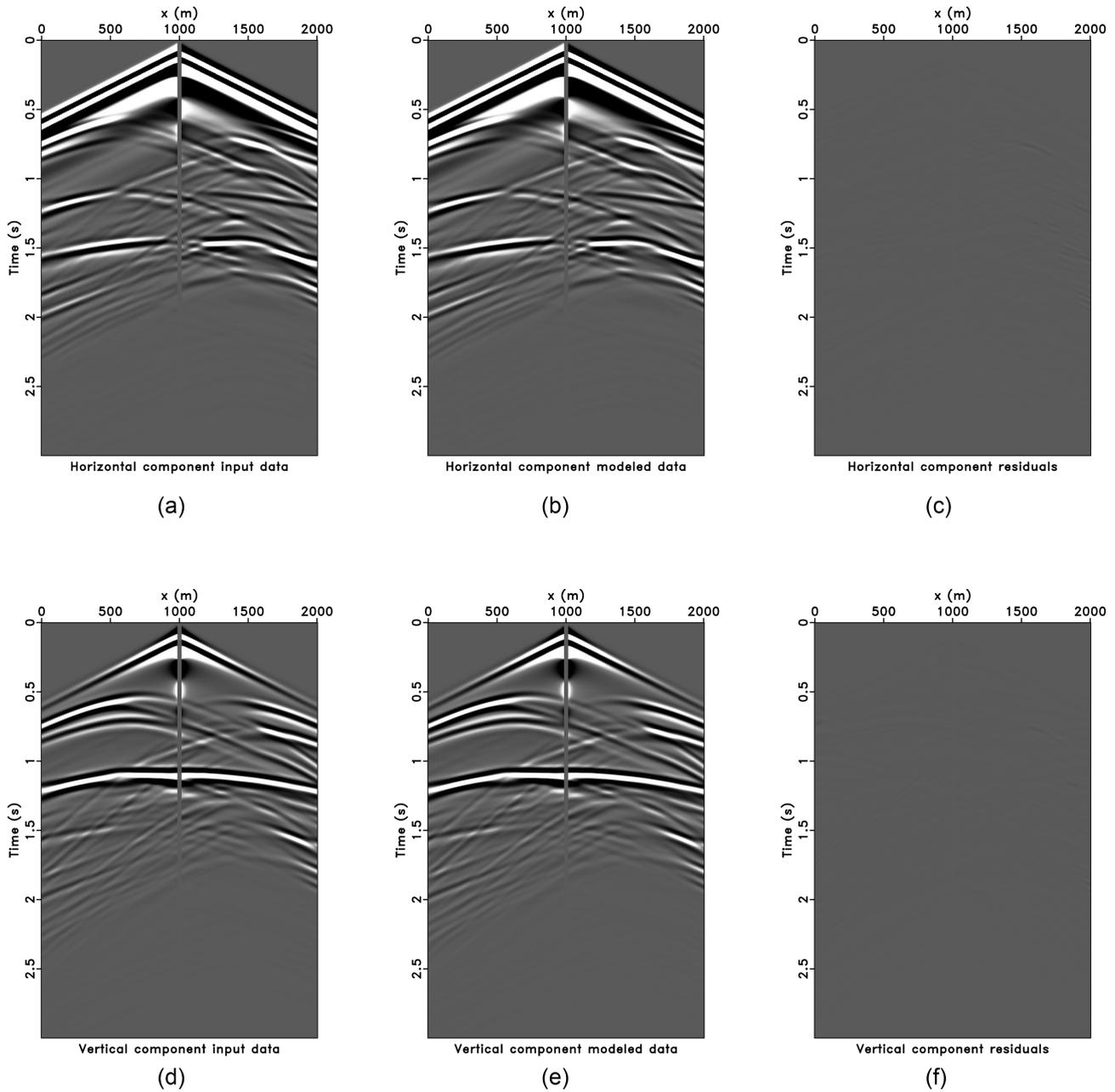


Figure 9. Data and residuals for elastic Gauss–Newton FWI. (a) Observed horizontal component data of shot at $x = 1000$ m. (b) Horizontal component data modelled by elastic Gauss–Newton FWI inverted models. (c) Horizontal component data residual. (d) Observed vertical component data of shot at $x = 1000$ m. (e) Vertical component data modelled by elastic Gauss–Newton FWI inverted models. (f) Vertical component data residual.

where $[\partial J(\mathbf{m})]/[\partial \mathbf{m}]$ is the gradient and $[\partial^2 J(\mathbf{m})]/[\partial \mathbf{m}^2]$ is the Hessian of the objective function. In the vicinity of \mathbf{m} , the objective function is linearized and an optimal model update $\delta \mathbf{m}$ should satisfy $\partial J(\mathbf{m} + \delta \mathbf{m})/\partial \delta \mathbf{m}^T = 0$. This leads to the following Newton step

$$\begin{aligned} \delta \mathbf{m} &= - \left[\frac{\partial^2 J(\mathbf{m})}{\partial \mathbf{m}^2} \right]^{-1} \frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} \\ &= -\mathbf{H}^{-1} \mathbf{g}, \end{aligned} \quad (11)$$

where $\mathbf{H} = [\partial^2 J(\mathbf{m})]/[\partial \mathbf{m}^2]$ denotes the Hessian operator and $\mathbf{g} = [\partial J(\mathbf{m})]/[\partial \mathbf{m}]$ denotes the gradient vector. Eq. (11) indicates that the gradient is preconditioned by the inverse of the Hessian. The inverse of Hessian acts as a refocusing, scaling and de-multiple operator (Pratt *et al.* 1998).

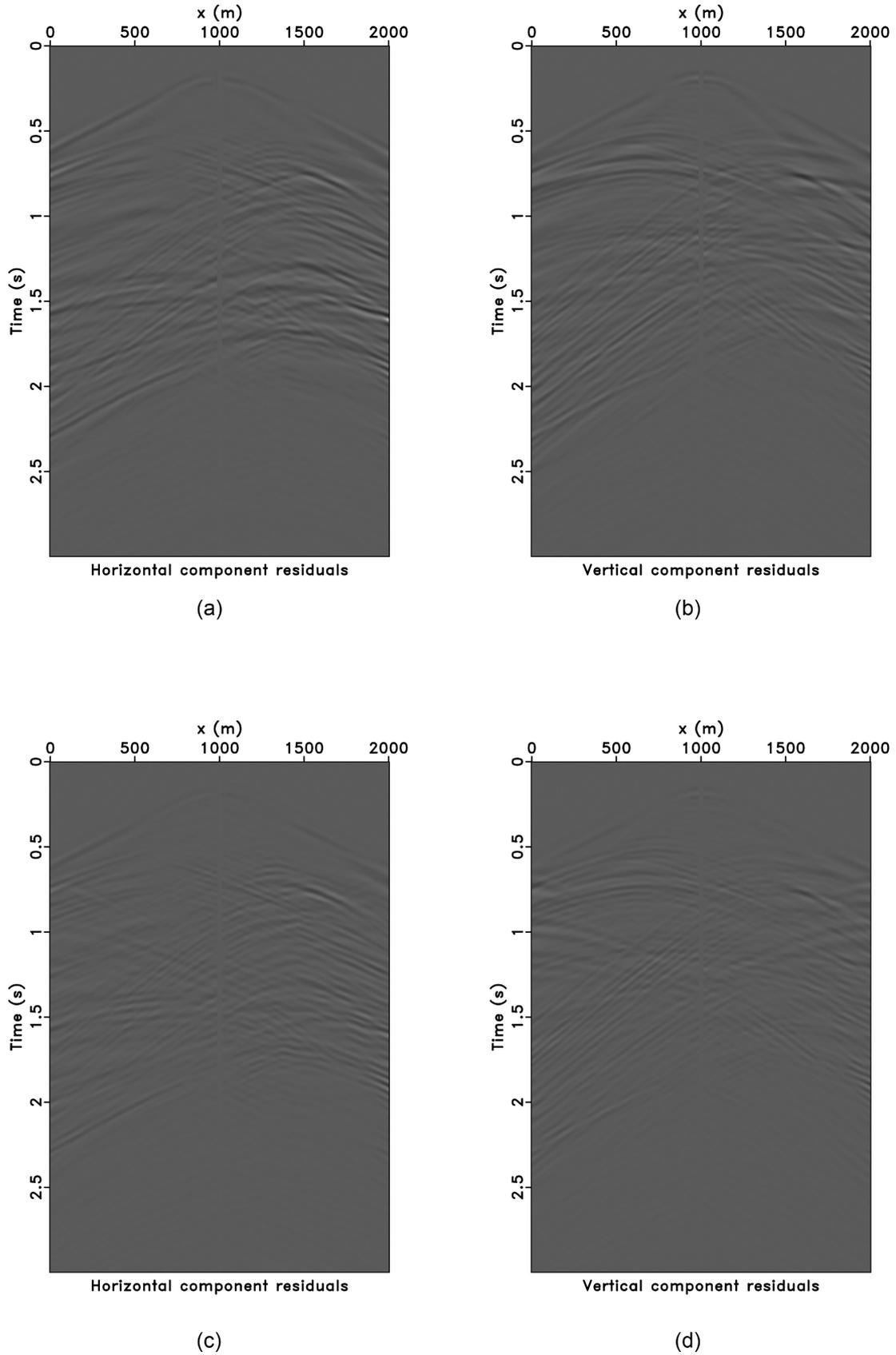
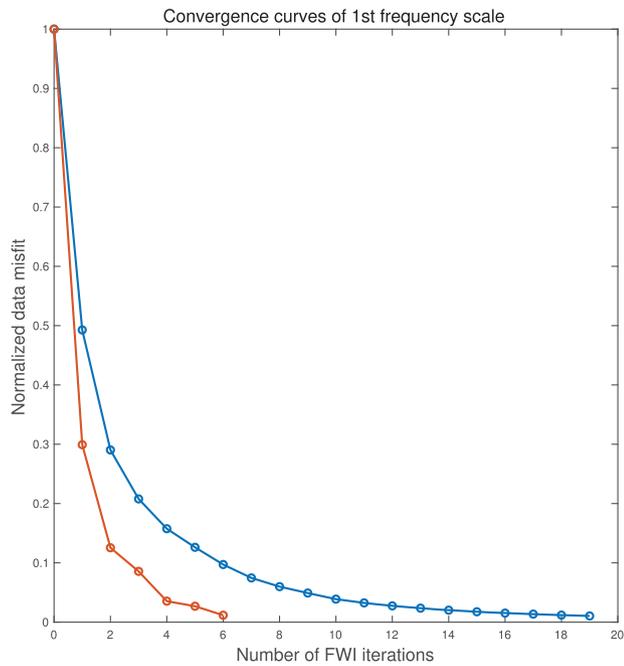
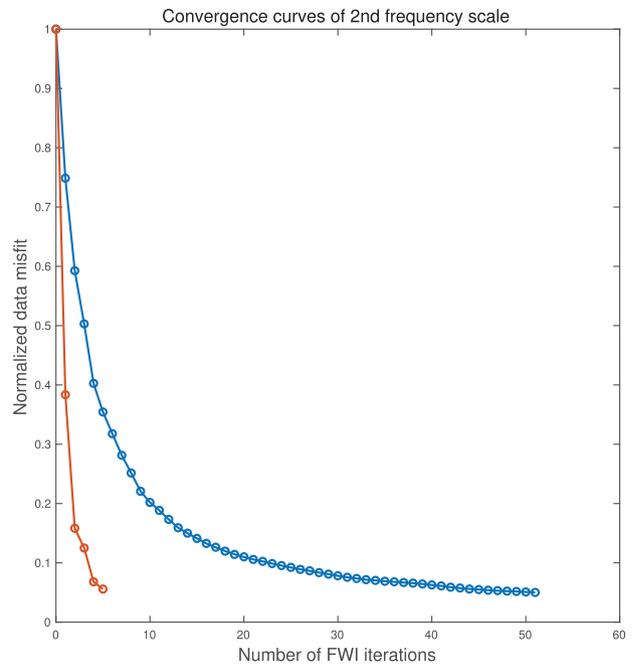


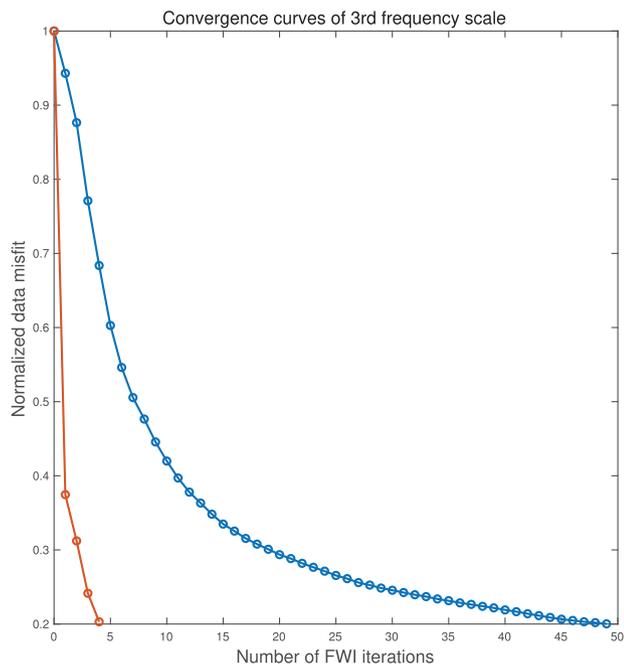
Figure 10. Data residuals comparison for shot at $x = 1000$ m. The data are plotted with heavy clipping (one fifth of the scale used in Figs 8 and 9). (a) Horizontal component data residual of elastic NLCG FWI. (b) Vertical component data residual of elastic NLCG FWI. (c) Horizontal component data residual of elastic Gauss–Newton FWI. (d) Vertical component data residual of elastic Gauss–Newton FWI.



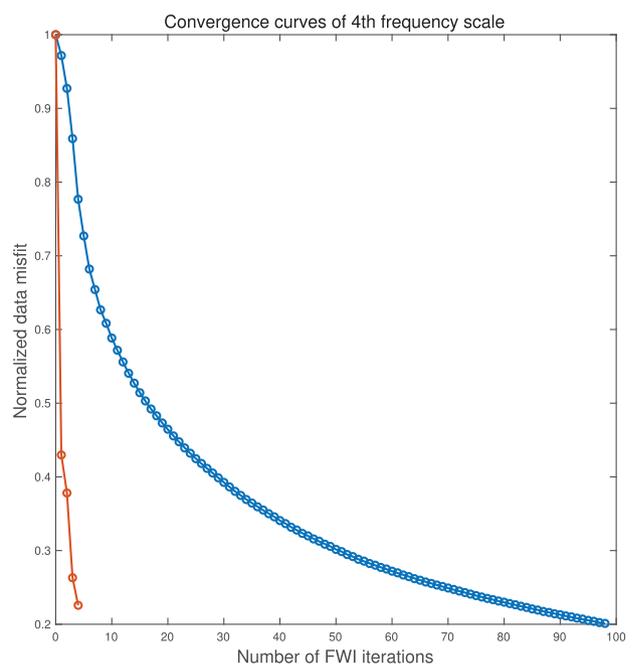
(a)



(b)



(c)



(d)

Figure 11. Normalized data misfit versus number of non-linear FWI iterations for elastic NLCG FWI and elastic Gauss–Newton FWI in frequency band (a) 3–5 Hz, (b) 3–10 Hz, (c) 3–15 Hz, (d) 3–30 Hz. Blue: data misfit curves for elastic NLCG FWI. Red: data misfit curves for elastic Gauss–Newton FWI.

The gradient is the derivative of the objective function with respect to model parameters

$$\mathbf{g} = \frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} = \sum_{i=1}^{N_s} \left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}} \right)^\dagger (\mathbf{d}_i - \mathbf{d}_i^{obs}), \quad (12)$$

where $\partial \mathbf{d}_i / \partial \mathbf{m}$ is the Fréchet derivative operator and the symbol \dagger denotes the adjoint of an operator. Eq. (12) indicates that the gradient of FWI is generated by applying the adjoint Fréchet derivative operator to data residuals. It is equivalent to a pre-stack elastic RTM algorithm applied to data residuals (Tarantola 1986).

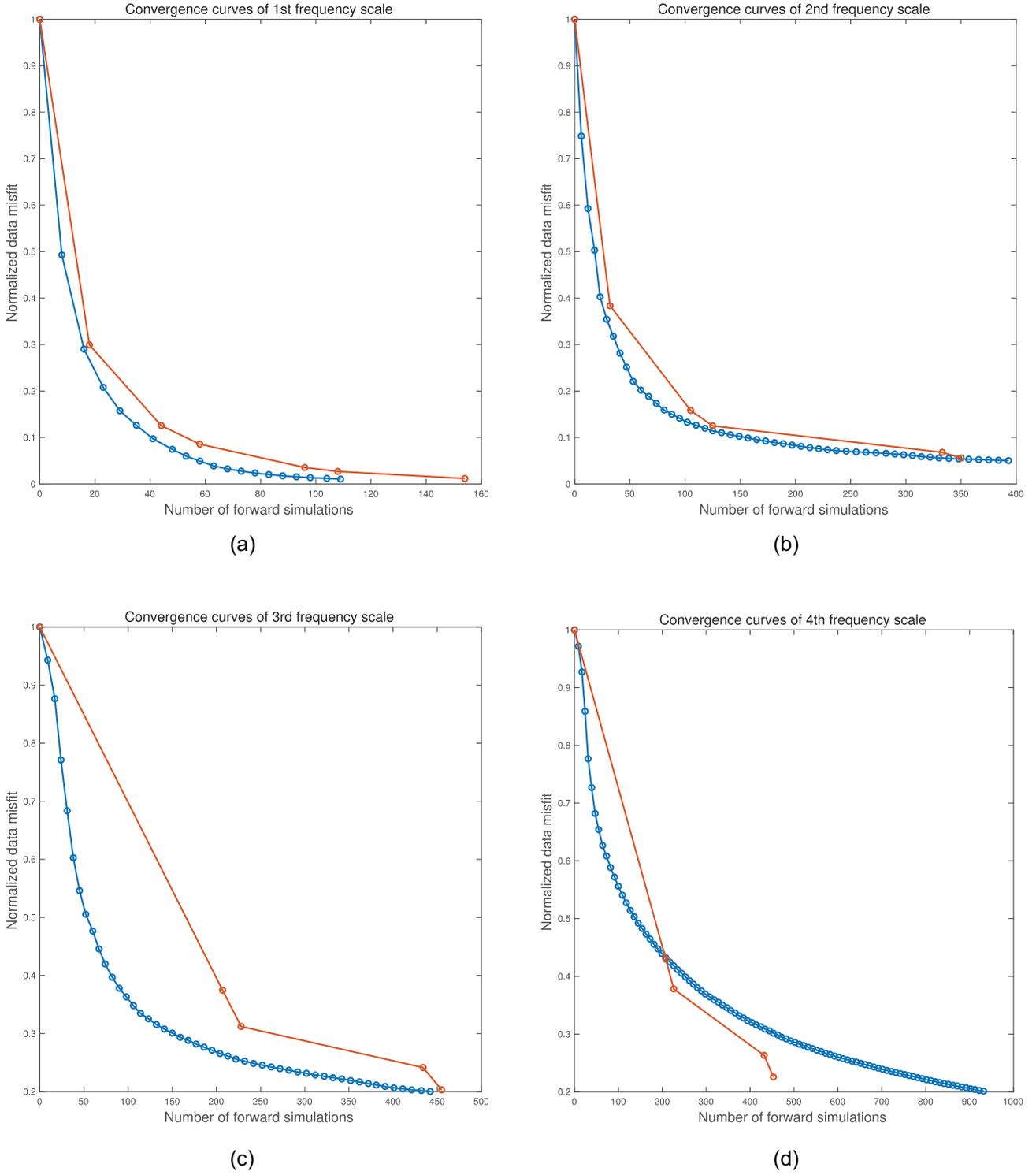


Figure 12. Normalized data misfit versus number of forward simulations for elastic NLCG FWI and elastic Gauss–Newton FWI in frequency band (a) 3–5 Hz, (b) 3–10 Hz, (c) 3–15 Hz, (d) 3–30 Hz. Blue: data misfit curves for elastic NLCG FWI. Red: data misfit curves for elastic Gauss–Newton FWI.

In a similar way, the Hessian is obtained by differentiating the gradient with respect to model parameters

$$\begin{aligned}
 \mathbf{H} &= \frac{\partial^2 J(\mathbf{m})}{\partial \mathbf{m}^2} \\
 &= \sum_{i=1}^{N_s} \left[\left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}} \right)^\dagger \left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}} \right) + \left(\frac{\partial^2 \mathbf{d}_i}{\partial \mathbf{m}^2} \right)^\dagger (\mathbf{d}_i - \mathbf{d}_i^{obs}) \right].
 \end{aligned}
 \tag{13}$$

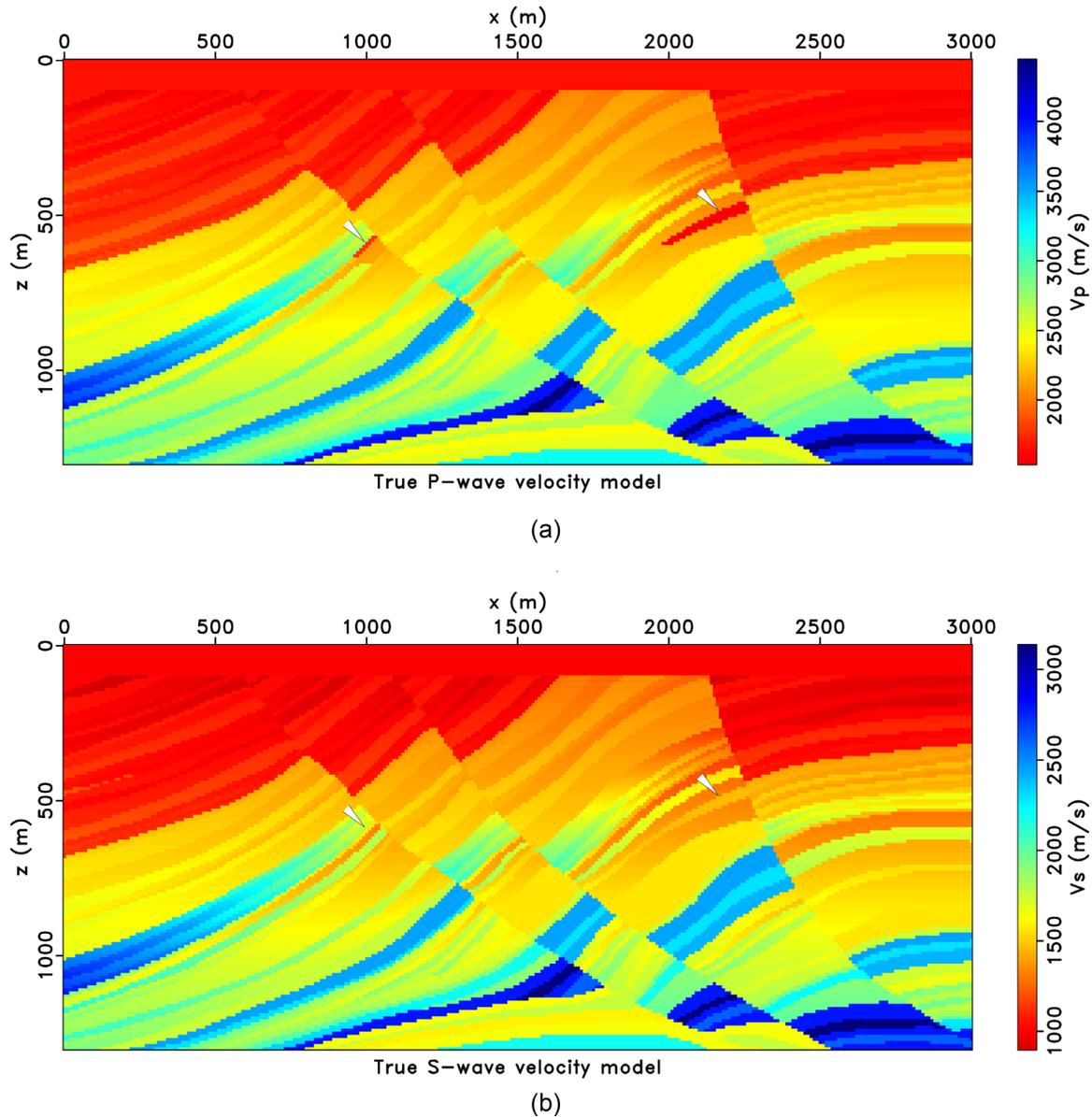


Figure 13. Elastic Marmousi2 model. (a) True P -wave velocity model. (b) True S -wave velocity model.

eq. (13) is the full Hessian of the FWI problem. The second-order partial derivative is related to second-order scattering effects. The second-order effect is small when the objective function is close to a minimum (Tarantola 2005). Dropping the second-order term in the full Hessian leads to the Gauss–Newton Hessian (Pratt *et al.* 1998; Fichtner & Trampert 2011)

$$\mathbf{H}_{GN} = \sum_{i=1}^{N_s} \left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}} \right)^\dagger \left(\frac{\partial \mathbf{d}_i}{\partial \mathbf{m}} \right). \quad (14)$$

The Gauss–Newton Hessian is the cascade of forward and adjoint Fréchet derivative operators. The Gauss–Newton step can be expressed as

$$\delta \mathbf{m} = -\mathbf{H}_{GN}^{-1} \mathbf{g}. \quad (15)$$

The inverse of Gauss–Newton Hessian focuses the gradient and compensates the uneven illumination of the subsurface (Pratt *et al.* 1998). The Gauss–Newton method for elastic FWI minimizes the non-linear least-squares objective function by iteratively updating the model. The updated model at $(k + 1)$ th iteration can be written as

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \eta^{(k)} \delta \mathbf{m}^{(k)} = \mathbf{m}^{(k)} - \eta^{(k)} (\mathbf{H}_{GN}^{(k)})^{-1} \mathbf{g}^{(k)}, \quad (16)$$

where $\mathbf{m}^{(k)}$ is the current model, $\eta^{(k)}$ is a step-size scalar, and $\delta \mathbf{m}^{(k)}$ is the Gauss–Newton search direction, $\mathbf{H}_{GN}^{(k)}$ is the Gauss–Newton Hessian and $\mathbf{g}^{(k)}$ is the gradient. In our work, the step size $\eta^{(k)}$ is estimated by a parabolic interpolation through three points $[\eta, J(\mathbf{m}^{(k)} + \eta \delta \mathbf{m}^{(k)})]$ (Vigh

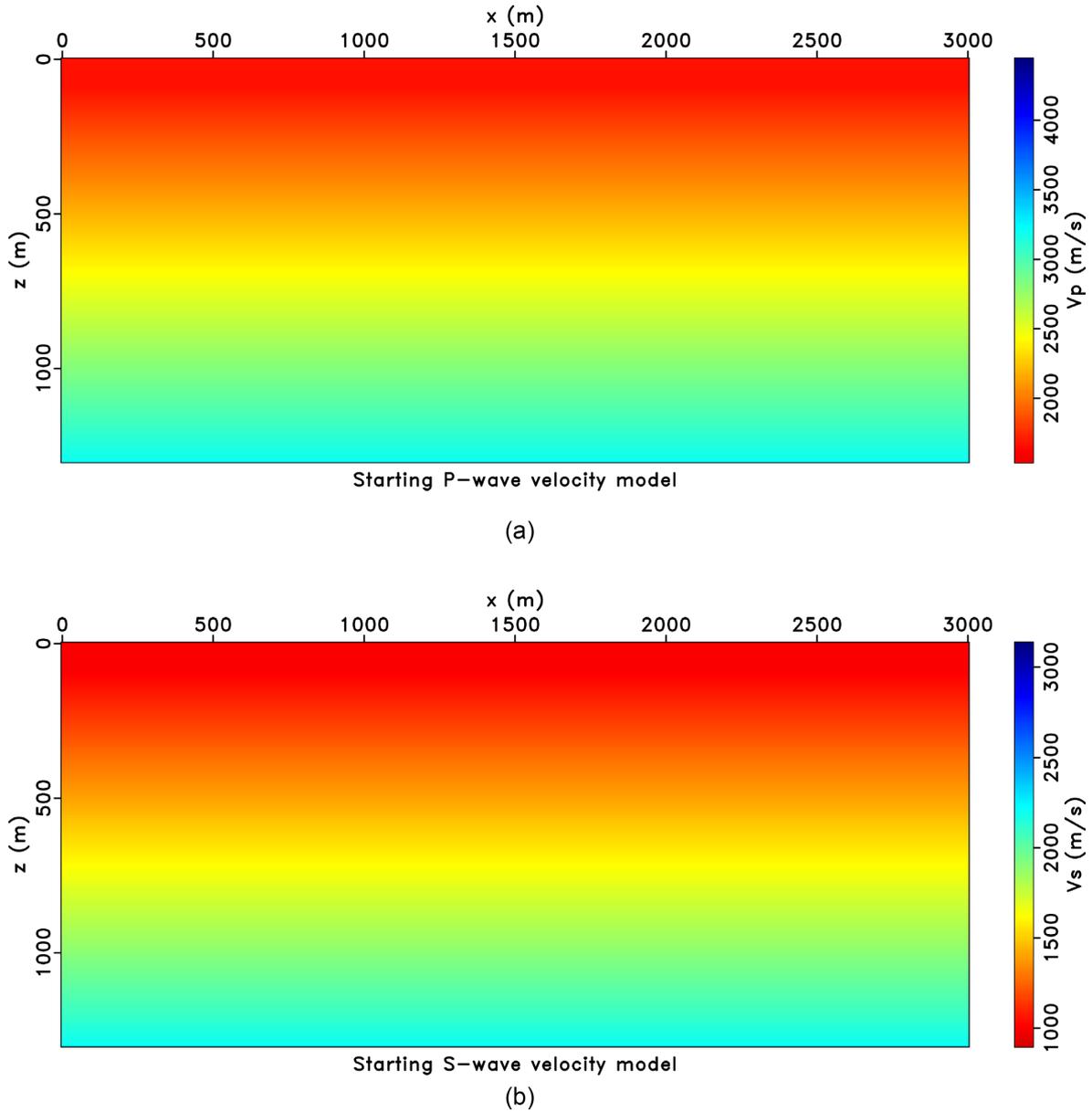


Figure 14. (a) Starting P -wave velocity model. (b) Starting S -wave velocity model.

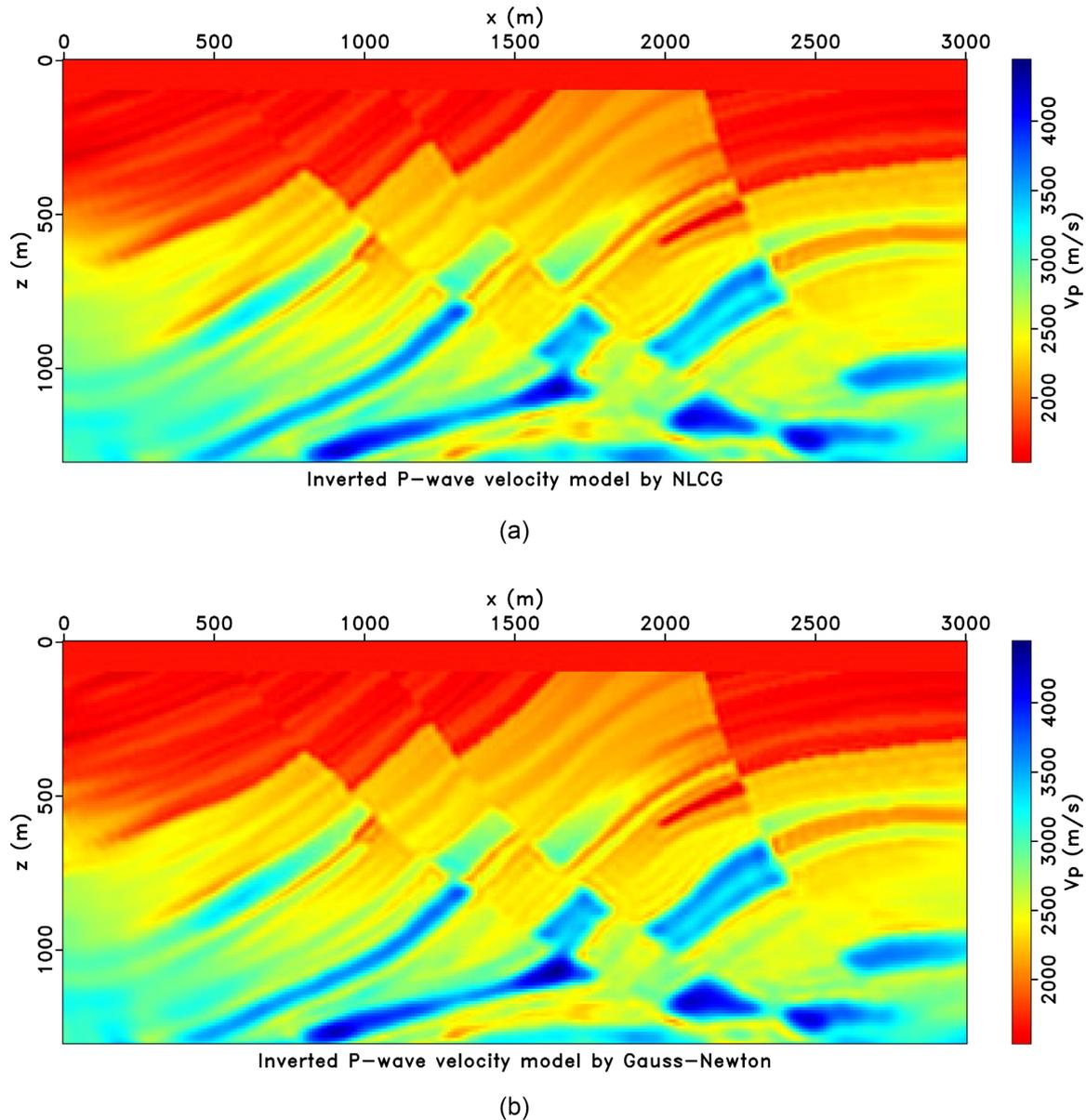
et al. 2009; Virieux & Operto 2009). The general formulation of the time-domain elastic Gauss–Newton FWI algorithm is summarized by Algorithm 1

Algorithm 1 Time-domain elastic Gauss–Newton FWI

Initialize: $\mathbf{m}^{(0)}$
for $k = 0, 1, \dots$ **while** it does not converge **do**
 1. Forward modelling: $\mathbf{d}_i^{(k)} = \mathbf{d}_i(\mathbf{m}^{(k)})$, $i = 1, \dots, N_s$
 2. Compute data residuals: $\delta \mathbf{d}_i^{(k)} = \mathbf{d}_i^{obs} - \mathbf{d}_i^{(k)}$, $i = 1, \dots, N_s$
 3. Compute Gauss–Newton search direction: $\delta \mathbf{m}^{(k)} = -(\mathbf{H}_{GN}^{(k)})^{-1} \mathbf{g}^{(k)}$
 4. Compute step size $\eta^{(k)}$ via parabolic fitting line search.
 5. $\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \eta^{(k)} \delta \mathbf{m}^{(k)}$
end

Table 2. Parameters used in the elastic Marmousi2 model inversion example.

Frequency group	Stopping criterion for FWI loop ϵ	Stopping criterion for CGLS loop ξ
2–3 Hz	0.01	0.5
2–5 Hz	0.1	0.5
2–10 Hz	0.1	0.5
2–30 Hz	0.1	0.5

**Figure 15.** (a) Inverted P -wave velocity model by elastic NLCG FWI. (b) Inverted P -wave velocity model by elastic Gauss–Newton FWI.

2.3 Solving for the Gauss–Newton step via the CGLS algorithm combined with the adjoint-state method

2.3.1 Formulation of the Gauss–Newton step as an elastic LSRTM problem

In this section, we will drop the FWI iteration index k from our mathematical expression to make the notation more concise but we understand that the quantities depend on the FWI iteration index. In the definition of the general Gauss–Newton iteration algorithm (eq. 16), the Gauss–Newton Hessian \mathbf{H}_{GN} needs to be inverted. However, explicitly computing, storing and inverting the Hessian are prohibitively expensive operations for realistic scale problem. Instead, we solve for the Gauss–Newton step update via a matrix-free conjugate gradient least-squares algorithm in conjunction with the adjoint-state method. The Gauss–Newton step 15 can be computed via solving the linear system of equations

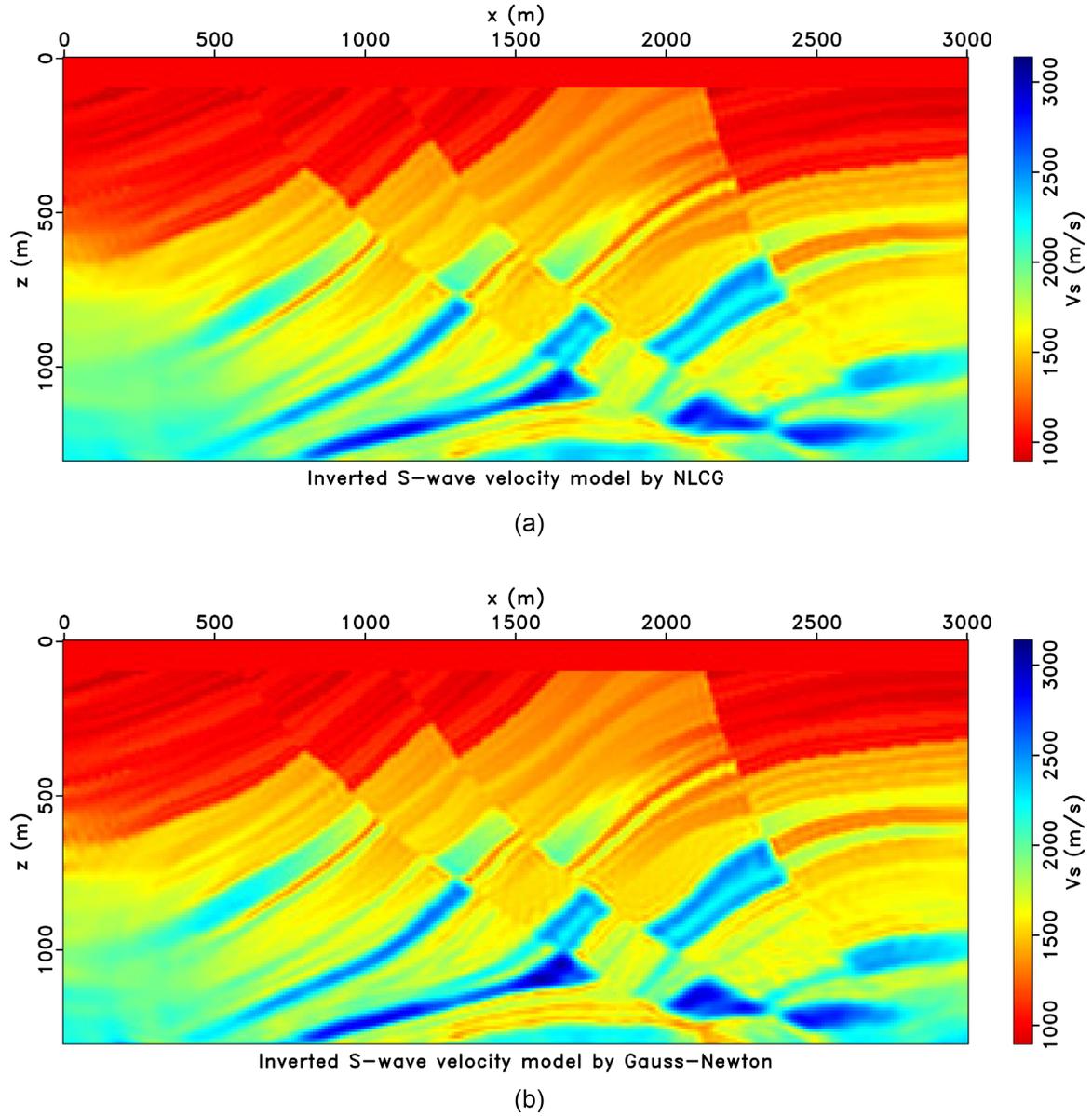


Figure 16. (a) Inverted S -wave velocity model by elastic NLCG FWI. (b) Inverted S -wave velocity model by elastic Gauss–Newton FWI.

given by

$$\mathbf{H}_{GN} \delta \mathbf{m} = -\mathbf{g}. \quad (17)$$

If we use \mathbf{L}_i to denote the Fréchet derivative operator ($\partial \mathbf{d}_i / \partial \mathbf{m}$) and $\delta \mathbf{d}_i$ to denote the data residual ($\mathbf{d}_i^{obs} - \mathbf{d}_i$) for the i th source, the linear system (17) becomes

$$\sum_{i=1}^{N_s} \mathbf{L}_i^\dagger \mathbf{L}_i \delta \mathbf{m} = \sum_{i=1}^{N_s} \mathbf{L}_i^\dagger \delta \mathbf{d}_i. \quad (18)$$

Solving the above system entails the solution of a least-squares inversion problem where we minimize the following cost function

$$J_{GN}(\delta \mathbf{m}) = \frac{1}{2} \sum_{i=1}^{N_s} \|\mathbf{L}_i \delta \mathbf{m} - \delta \mathbf{d}_i\|_2^2. \quad (19)$$

We recognize that this is actually the elastic LSRTM formulation (Chen & Sacchi 2017, 2019). If we adopt the exploration imaging jargon, the Fréchet derivative operator \mathbf{L}_i is indeed the elastic Born modelling operator, and the adjoint of Fréchet derivative operator \mathbf{L}_i^\dagger is the elastic RTM operator. We choose to solve the optimization problem (19) using the conjugate gradient least-squares algorithm that entails the implicit-form application of the elastic Born modelling operator and the elastic RTM operator. In the appendix, we present the derivations

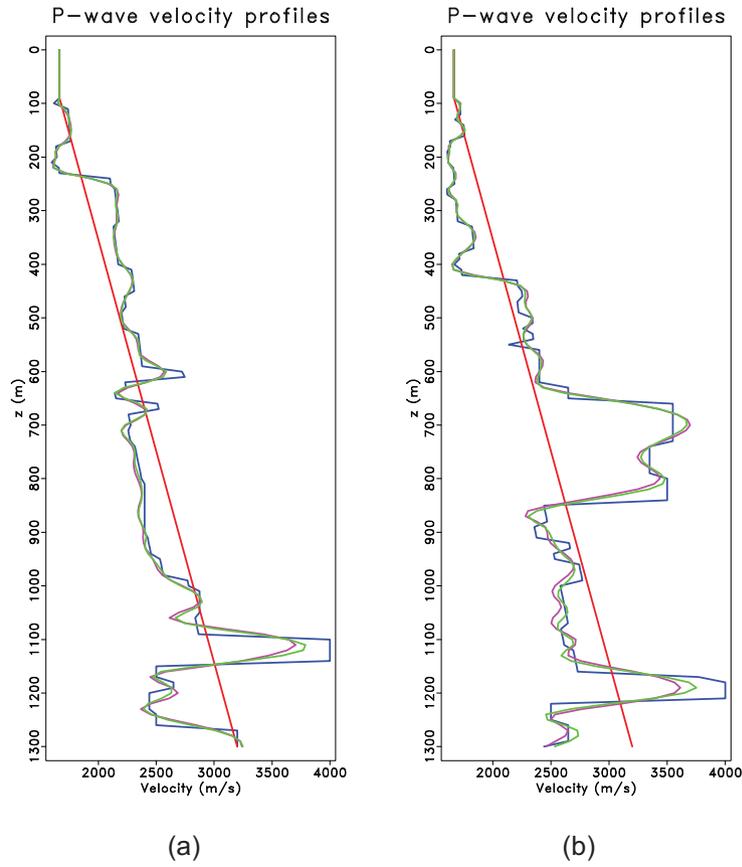


Figure 17. Profiles of inverted P -wave velocity models at (a) $x = 1500$ m and (b) $x = 2300$ m. Blue: true model; Red: starting model; Purple: inverted model by elastic NLCG FWI; Green: inverted model by elastic Gauss-Newton FWI.

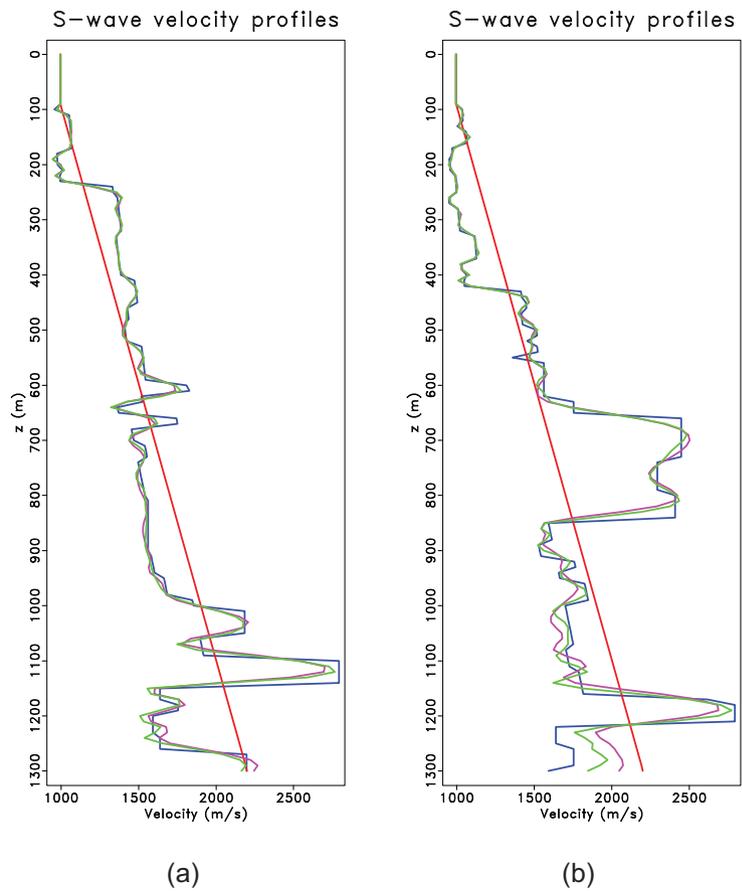


Figure 18. Profiles of inverted S -wave velocity models at (a) $x = 1500$ m and (b) $x = 2300$ m. Blue: true model; Red: starting model; Purple: inverted model by elastic NLCG FWI; Green: inverted model by elastic Gauss-Newton FWI.

of the operators \mathbf{L}_i and \mathbf{L}_i^\dagger for the 3-D isotropic elastic case which are the expressions for the elastic Born approximation modelling and the elastic RTM kernel.

2.3.2 Solving for the Gauss–Newton step using preconditioned CGLS algorithm

It is important to precondition the linear system of equations (eq. 18) to accelerate the convergence of the CGLS algorithm. For this purpose, we adopt the elastic pseudo-Hessian to derive a pre-conditioning operator (Shin *et al.* 2001; Chen & Sacchi 2017). The preconditioned version of the Gauss–Newton step minimizes

$$J_{GN}(\delta\tilde{\mathbf{m}}) = \frac{1}{2} \sum_{i=1}^{N_s} \|\mathbf{L}_i \mathbf{P} \delta\tilde{\mathbf{m}} - \delta\mathbf{d}_i\|_2^2, \quad (20)$$

where \mathbf{P} represents the pseudo-Hessian preconditioning operator. The preconditioned conjugate gradient least squares (PCGLS) algorithm (Bjorck 1996) is summarized in Algorithm 2.

Algorithm 2 Preconditioned CGLS algorithm

```

Initialize
 $\delta\mathbf{m}^{(0)} = \mathbf{0}$ 
 $\mathbf{r}_i^{(0)} = \delta\mathbf{d}_i, i = 1, \dots, N_s$ 
 $\mathbf{s}^{(0)} = \mathbf{P}^\dagger \left( \sum_{i=1}^{N_s} \mathbf{L}_i^\dagger \mathbf{r}_i^{(0)} \right)$  //compute the preconditioned gradient
 $\mathbf{c}^{(0)} = \mathbf{s}^{(0)}$ 
 $\gamma^{(0)} = \|\mathbf{s}^{(0)}\|_2^2$ 
for  $l = 0, 1, \dots$  while not converge do
 $\mathbf{t}^{(l)} = \mathbf{P}\mathbf{c}^{(l)}$ ,
 $\mathbf{q}_i^{(l)} = \mathbf{L}_i \mathbf{t}^{(l)}, i = 1, \dots, N_s$  //Born forward modelling
 $\delta^{(l)} = \sum_{i=1}^{N_s} \|\mathbf{q}_i^{(l)}\|_2^2$ 
 $\alpha^{(l)} = \gamma^{(l)} / \delta^{(l)}$  //calculate the step size
 $\delta\mathbf{m}^{(l+1)} = \delta\mathbf{m}^{(l)} + \alpha^{(l)} \mathbf{t}^{(l)}$  //update the model
 $\mathbf{r}_i^{(l+1)} = \mathbf{r}_i^{(l)} - \alpha^{(l)} \mathbf{q}_i^{(l)}, i = 1, \dots, N_s$  //compute data residuals
 $\mathbf{s}^{(l+1)} = \mathbf{P}^\dagger \left( \sum_{i=1}^{N_s} \mathbf{L}_i^\dagger \mathbf{r}_i^{(l+1)} \right)$  //compute the preconditioned gradient
 $\gamma^{(l+1)} = \|\mathbf{s}^{(l+1)}\|_2^2$ 
 $\beta^{(l)} = \gamma^{(l+1)} / \gamma^{(l)}$ 
 $\mathbf{c}^{(l+1)} = \mathbf{s}^{(l+1)} + \beta^{(l)} \mathbf{c}^{(l)}$  //compute the conjugate direction
end

```

The output of Algorithm 2 is the inverted multiparameter model perturbation $\delta\mathbf{m} = \mathbf{P}\delta\tilde{\mathbf{m}}$. The estimated model perturbation is also the Gauss–Newton search direction.

2.4 Time-domain matrix-free elastic Gauss–Newton FWI

As discussed in the previous section, we propose a time-domain elastic Gauss–Newton FWI using CGLS algorithm to solve the internal Gauss–Newton linear system of equations. We stress that one does not need to formulate or invert the Hessian matrix explicitly. Instead, the solution is iteratively retrieved by the CGLS method. The CGLS only requires two operators \mathbf{L} (Fréchet derivative) and \mathbf{L}^\dagger (the adjoint Fréchet derivative) that are applied in implicit-form to vectors. The operators \mathbf{L} and \mathbf{L}^\dagger are efficiently applied to vectors via the adjoint-state method. We name this algorithm, the *time-domain matrix-free elastic Gauss–Newton FWI*. The algorithm is compiled in Algorithm 3. In step 3 of Algorithm 3, Algorithm 2 is adopted to solve for the Gauss–Newton search direction.

3 NUMERICAL EXAMPLES

3.1 Source code main features

Our computer code is developed using the C language and parallelized via the Message Passing Interface (MPI). Our forward modelling code adopts a time-domain staggered-grid finite-difference scheme (Virieux 1986) to discretize the elastic wave equation. The algorithm uses the unsplit convolutional perfectly matched layer (C-PML) boundary condition (Komatitsch & Martin 2007) to absorb the artificial reflections arising from the computational boundaries. In our algorithm, the order of finite-difference in space is selectable. The code automatically computes the finite-difference coefficients from the user-specified finite-difference order (Liu & Sen 2009). When computing the action of

Algorithm 3 Time-domain matrix-free elastic Gauss-Newton FWI**Initialize:** $\mathbf{m}^{(0)}$ **for** $k = 0, 1, \dots$ **while** it does not converge **do**

1. Forward modelling: $\mathbf{d}_i^{(k)} = \mathbf{d}_i(\mathbf{m}^{(k)})$, $i = 1, \dots, N_s$
2. Compute data residual: $\delta \mathbf{d}_i^{(k)} = \mathbf{d}_i^{obs} - \mathbf{d}_i^{(k)}$, $i = 1, \dots, N_s$
3. Compute Gauss-Newton search direction $\delta \mathbf{m}^{(k)}$ by solving

$$\min_{\delta \mathbf{m}^{(k)}} \frac{1}{2} \sum_{i=1}^{N_s} \|\mathbf{L}_i^{(k)} \mathbf{P}^{(k)} \delta \mathbf{m}^{(k)} - \delta \mathbf{d}_i^{(k)}\|_2^2$$

Apply the preconditioned CGLS algorithm 2.

4. Compute step size $\eta^{(k)}$ via parabolic fitting line search.

$$5. \mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \eta^{(k)} \delta \mathbf{m}^{(k)}$$

end

the adjoint Fréchet derivative operator on a vector (eq. A21), we do not save the entire source-side wavefield. Instead, we adopt the source-wavefield reconstruction method (Gauthier *et al.* 1986; Dussaud *et al.* 2008) that allows us to only save wavefields on boundaries within the depth of half of the spatial finite-difference operator length and the final time frame during the forward simulation of state equation. Then, the state variable \mathbf{u} (source-side wavefield) is recomputed from the saved wavefield by backward propagation while simultaneously computing the adjoint variable \mathbf{p} (receiver-side wavefield). This method is valid because the elastic wave equation is self-adjoint and reversible (Tarantola 1988). Our code for the Fréchet derivative operator (\mathbf{L} , elastic Born operator) and adjoint Fréchet operator (\mathbf{L}^\dagger , elastic RTM operator) pass the dot-product test (Mora 1987b; Claerbout 1992) which guarantees that these operators act as matrix-vector and transpose matrix-vector multiplications to ensure the convergence of the conjugate gradient least-squares method. The inversion follows Algorithm 3 and Algorithm 2. Also note that, our elastic Gauss–Newton FWI utilizes a multiscale strategy (Kolb *et al.* 1986; Bunks *et al.* 1995). The time-domain data are filtered to form sets of data with different frequency bands. Our FWI code inverts the low-frequency band data first and then inverts high-frequency data. The inverted models from an early scale are used as the initial model for inversion of higher scales. That is to say, there is another loop over frequency bands outside the iterations in Algorithm 3. There are in total three iterative loops: the outermost loop for iterating over different frequency bands, the loop for the Gauss–Newton FWI model update, and the innermost iteration of the CGLS method to estimate the Gauss–Newton direction.

3.2 Stopping criterion

As mentioned above, our time-domain inversion adopts the multiscale method. The inversion starts from filtered low-frequency data and gradually expands the frequency band (i.e. f_1 – f_2 Hz, f_1 – f_3 Hz, f_1 – f_4 Hz, \dots). The data misfit for a particular frequency band is defined as

$$J(\mathbf{m}^{(k)}) = \sum_{i=1}^{N_s} \|\mathbf{d}_i(\mathbf{m}^{(k)}) - \mathbf{d}_i^{obs}\|_2^2, \quad (21)$$

where k is the number of FWI iterations. The stopping criterion for the outer FWI loop for a frequency band is based on the normalized data misfit in that particular frequency band

$$\frac{J(\mathbf{m}^{(k)})}{J(\mathbf{m}^{(0)})} < \epsilon. \quad (22)$$

The stopping criterion ϵ is chosen differently for different frequency band because the inversion converges faster for initial frequency bands. The stopping criterion for the inner loop in the k th outer FWI iteration is given by

$$\frac{\|\mathbf{H}_{GN}^{(k)} \delta \mathbf{m}^{(k)} + \mathbf{g}^{(k)}\|_2}{\|\mathbf{g}^{(k)}\|_2} < \xi. \quad (23)$$

The inner minimization loop is stopped when the relative change in the gradient falls below a tolerance (Gratton *et al.* 2007). In other words, the inner loop of the Gauss–Newton method is truncated and solved inexactly. More sophisticated stopping schemes such as the Eisenstat–Walker forcing-term method can also be adopted (Eisenstat & Walker 1996; Epanomeritakis *et al.* 2008; Metivier *et al.* 2013).

The time-domain matrix-free elastic FWI method is tested on an elastic inclusion model and the elastic Marmousi2 model (Martin *et al.* 2006). The time-domain elastic staggered-grid finite-difference approach is adopted to synthesize observations with a C-PML boundary condition applied to the four boundaries of the model. In other words, we do not investigate the inversion of surface waves. The observed data are two-component particle velocity fields. In this study, we simultaneously invert for P - and S -wave velocities and assume that the density is known. Last, we also point out that the finite-difference operator adopted to synthesize the data and the operators utilized for the inversion shared the same grid. They also describe the same propagation phenomena, which is the frequently accepted manner to test FWI algorithms.

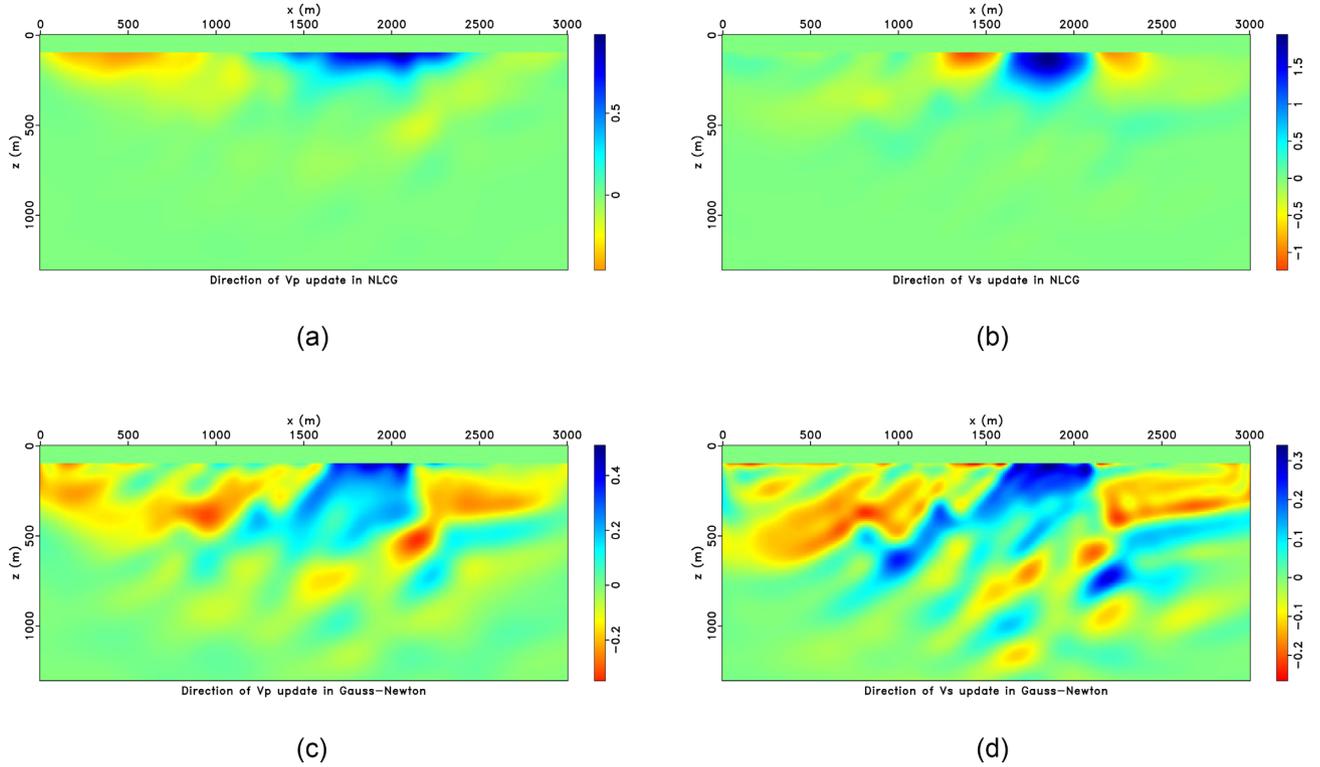


Figure 19. The model update direction in the first non-linear FWI iteration in the first frequency band (2–3 Hz). (a) Direction of V_p update in elastic NLCG FWI (negative of gradient direction). (b) Direction of V_s update in elastic NLCG FWI. (c) Direction of V_p update in elastic Gauss–Newton FWI (solved by CGLS). (d) Direction of V_s update in elastic Gauss–Newton FWI. Figures are plotted with original scale.

3.3 Elastic inclusion model

We adopt an elastic inclusion model to analyse the resolution of the proposed elastic Gauss–Newton FWI. Fig. 1 shows true P - and S -wave velocity models. Square shape velocity anomalies are embedded in a two-layer model at different locations. The subsurface density is fixed at 2000 kg m^{-3} . The model has a dimension of 2.0 km in the horizontal axis and 1.5 km depth with 201×151 gridpoints. There are in total 51 shots and 201 receivers deployed along the surface that simulates a fixed-spread survey geometry. The shot interval is 40 m, and receiver interval is 10 m. A 10 Hz central frequency Ricker wavelet is used to simulate an explosive source. The observed multicomponent data are simulated using our time-domain elastic staggered-grid finite-difference code. In this example, the forward modelling scheme adopts second-order finite differences in time and eighth-order finite differences in space ($\mathcal{O}(\Delta t^2, \Delta x^8)$). The 0–3 Hz frequency content is removed from the data using a minimum phase Butterworth filter to create a realistic inversion of data with poor content of low frequencies. We also mute near-offset traces in the data to mimic a more realistic scenario. We compare the results of conventional elastic FWI based on the non-linear conjugate gradient method (hereafter, we call it elastic NLCG FWI for short) and the proposed elastic Gauss–Newton FWI. In the elastic NLCG FWI, we used the elastic pseudo-Hessian operator to precondition the gradient. The line search is also based on parabolic fitting. The critical difference between the elastic NLCG FWI and elastic Gauss–Newton FWI can be summarized as follows. The model update in the elastic NLCG FWI is obtained by directly computing from the adjoint operator while the model update in the elastic Gauss–Newton FWI is obtained by solving a least-squares inverse problem (eq. 19). Both the elastic NLCG FWI and the elastic Gauss–Newton FWI utilize the multiscale method (Bunks *et al.* 1995). The four frequency bands for the multiscale inversion are 3–5 Hz, 3–10 Hz, 3–15 Hz and 3–30 Hz. Fig. 2 shows the starting P - and S -wave velocity models for elastic NLCG FWI and elastic Gauss–Newton FWI. The initial models contain two layers without the velocity anomalies.

For both the elastic NLCG FWI and the elastic Gauss–Newton FWI methods, we set the stopping criterion of the FWI loop to $\epsilon = 0.01, 0.05, 0.2, 0.2$ for the adopted four frequency bands, respectively. In other words, the elastic NLCG FWI and the elastic Gauss–Newton FWI converge to the same level of data misfit. This allows a fair comparison of the computational cost of the two methods. The stopping criterion for the inner loop of the Gauss–Newton method is set to be $\xi = 0.5$. The maximum number of CGLS iteration for the inner loop is set as 50. The parameters used in the inversion are summarized in Table 1. The inner minimization loop is not solved exactly to save computational cost. This strategy is often referred to as the truncated Gauss–Newton method (Gratton *et al.* 2007). The results of elastic NLCG FWI are shown in Figs 3(a) and 4(a). The results of elastic Gauss–Newton FWI are shown in Figs 3(b) and 4(b). Both elastic NLCG FWI and the proposed elastic Gauss–Newton obtain satisfactory results. The elastic Gauss–Newton FWI provides slightly better results than the elastic NLCG FWI. The elastic Gauss–Newton FWI results contain fewer artefacts and crosstalk contamination. To look at the details more closely, we display the profiles of the inverted P -wave velocity models in Fig. 5 and the profiles of the inverted S -wave velocity models in Fig. 6.

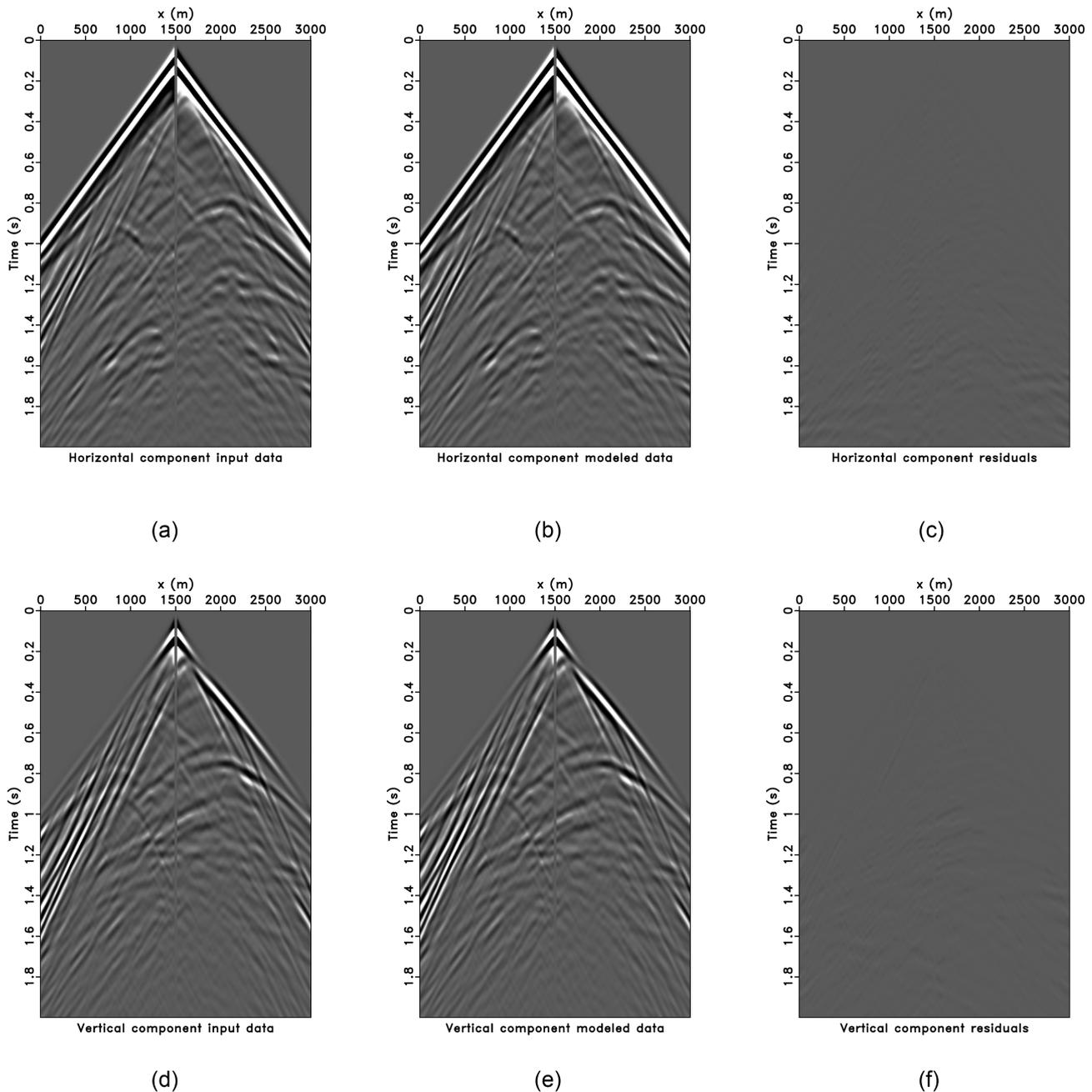


Figure 20. Data and residuals for elastic NLCG FWI. (a) Observed horizontal component data of shot at $x = 1500$ m. (b) Horizontal component data modelled by elastic NLCG FWI inverted models. (c) Horizontal component data residual. (d) Observed vertical component data of shot at $x = 1500$ m. (e) Vertical component data modelled by elastic NLCG FWI inverted models. (f) Vertical component data residual.

The reconstructed models by the elastic Gauss–Newton FWI are closer to the actual models than the ones reconstructed by the elastic NLCG FWI. In Fig. 7, we plot the model update direction of the two methods corresponding to first non-linear FWI iteration of the first frequency band (3–5 Hz). The Gauss–Newton method generates less crosstalk and artefacts in the V_p and V_s model update directions than the NLCG method.

We also evaluate data fitting for the models recovered by the two elastic FWI algorithms in Figs 8 and 9. Note that all the figures in Figs 8 and 9 are plotted using the same colour scale. The models obtained by both methods predict the data well. To have a closer look, we plot the data residuals with heavy clipping in Fig. 10 (one fifth of the scale used in Figs 8 and 9). From this figure, we can observe that the elastic Gauss–Newton FWI predicts the data slightly better than the elastic NLCG FWI.

Fig. 11 compares the convergence curves of normalized data misfit versus the number of non-linear iterations of the two methods for four different frequency bands. In terms of the number of non-linear FWI iterations, the elastic Gauss–Newton converges faster than the elastic NLCG FWI. One explanation is that the model search direction in the Gauss–Newton algorithm is more properly scaled than the one in the

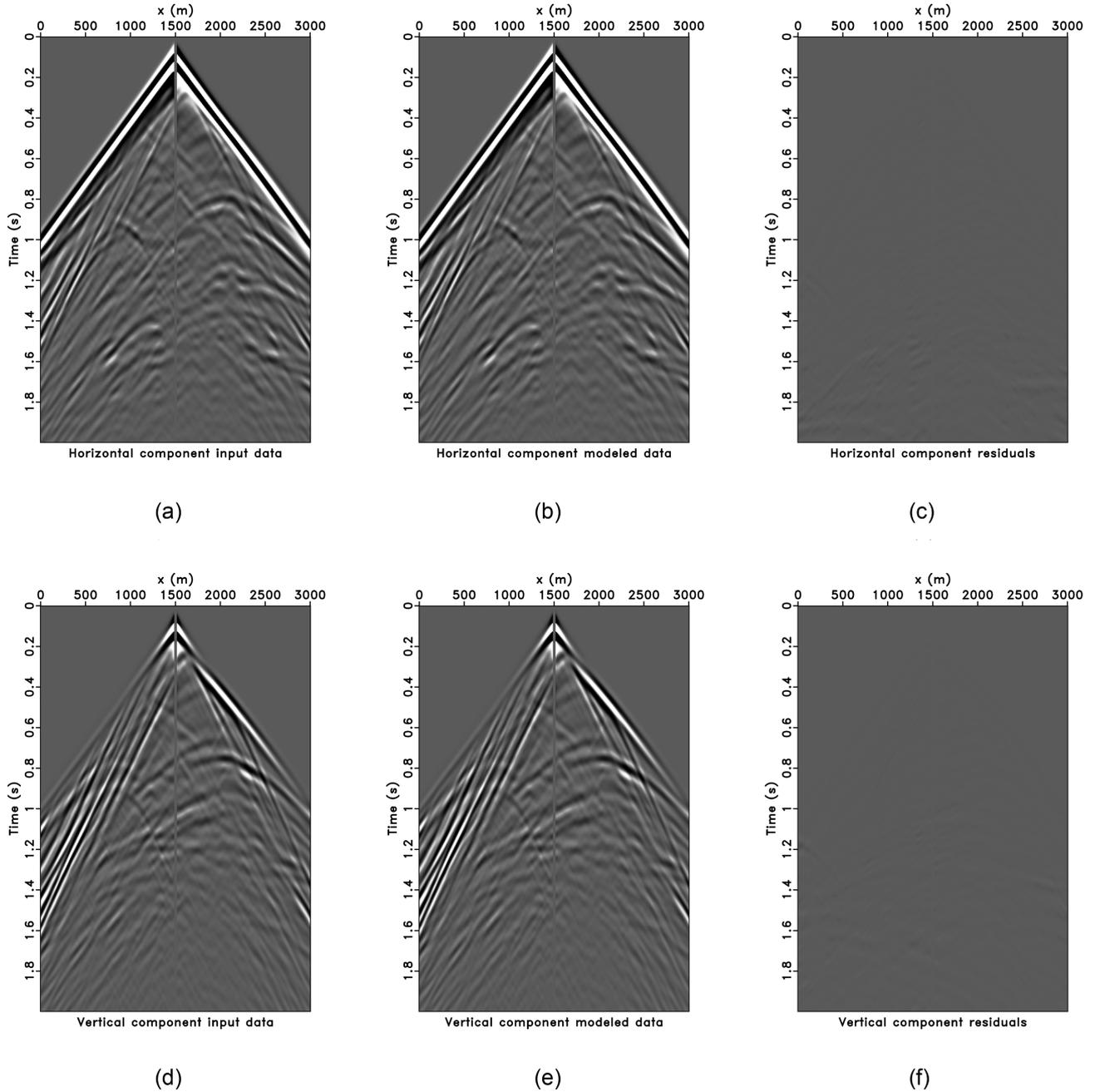


Figure 21. Data and residuals for elastic Gauss–Newton FWI. (a) Observed horizontal component data of shot at $x = 1500$ m. (b) Horizontal component data modelled by elastic Gauss–Newton FWI inverted models. (c) Horizontal component data residual. (d) Observed vertical component data of shot at $x = 1500$ m. (e) Vertical component data modelled by elastic Gauss–Newton FWI inverted models. (f) Vertical component data residual.

NLCG algorithm. The Gauss–Newton method has an additional iterative loop than the NLCG method. Thus, it is clear that the Gauss–Newton method requires substantially more computational effort per FWI iteration than NLCG FWI. We compare the computational cost of the two methods by plotting curves of the normalized data misfit versus the number of forward simulations performed of the two methods for four different frequency bands (Fig. 12). In this example, the total number of forward simulations performed by elastic NLCG FWI and elastic Gauss–Newton FWI is 1876 and 1412, respectively. In this elastic inclusion model inversion and under the above configuration, the elastic Gauss–Newton FWI is computationally more economical than the elastic NLCG FWI to converge to the same misfit level.

3.4 Elastic Marmousi2 model

Fig. 13 shows true P - and S -wave velocities of the elastic Marmousi2 model. The P - and S -wave velocity models are uncorrelated in the hydrocarbon reservoir areas that were indicated by the white triangles. Those areas are identified by low P -wave velocities, whereas their signature is much weaker in the S -wave velocity model. Density is assumed to be constant (2000 kg m^{-3}). The model has a dimension of

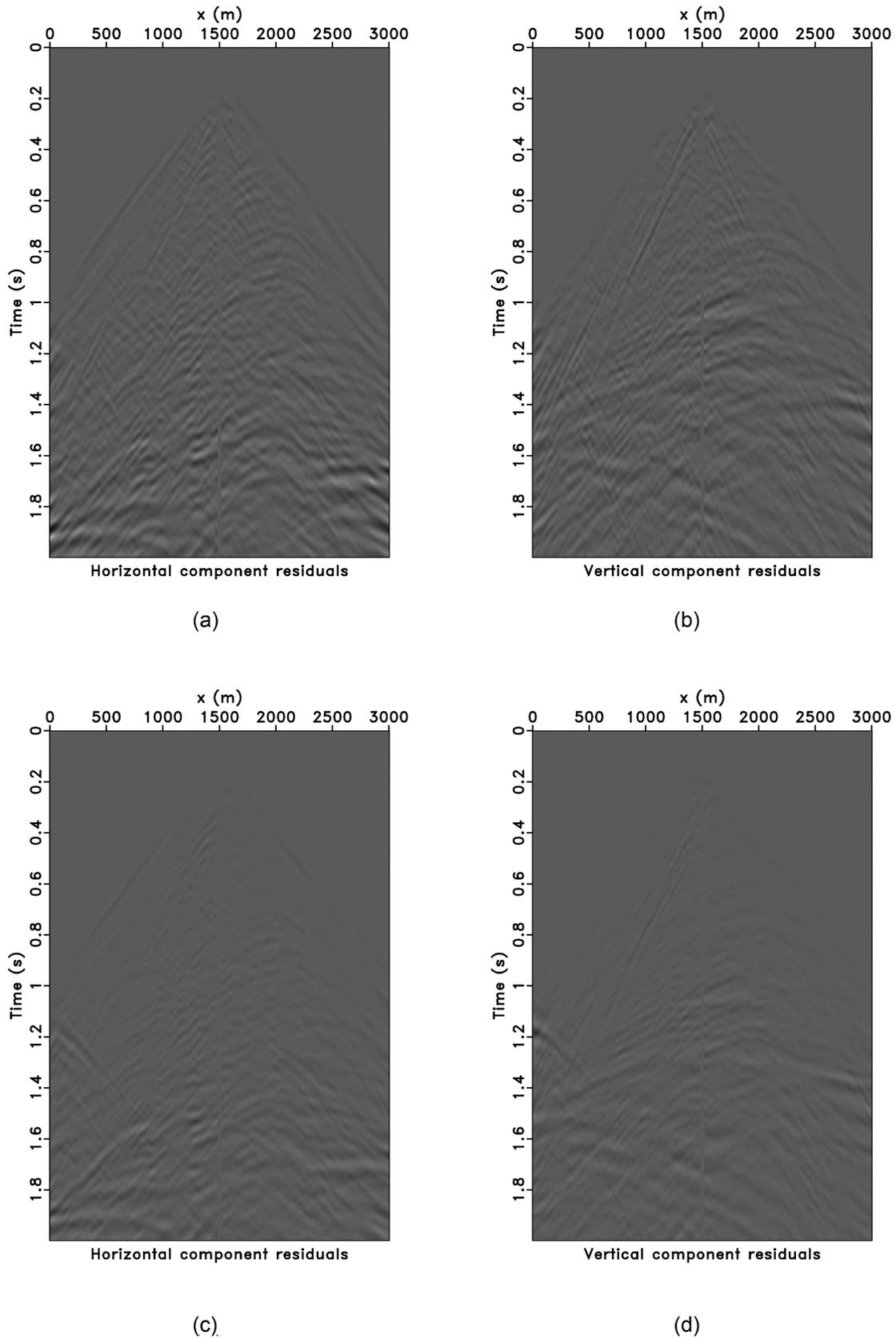
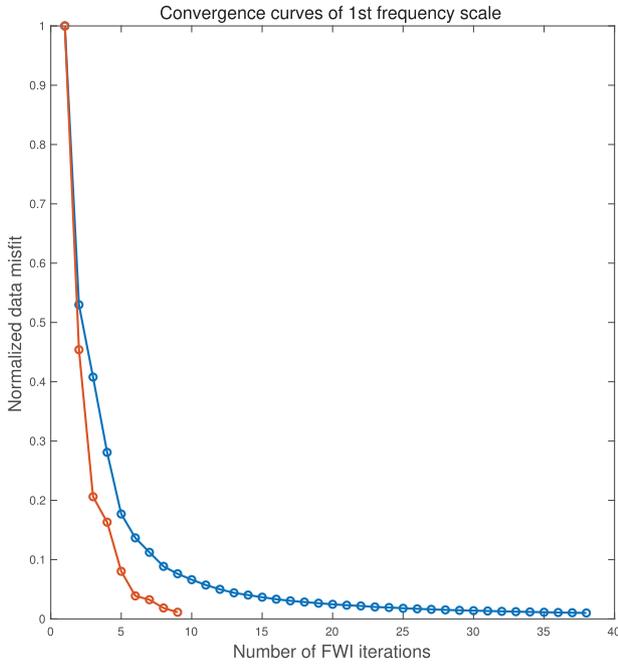
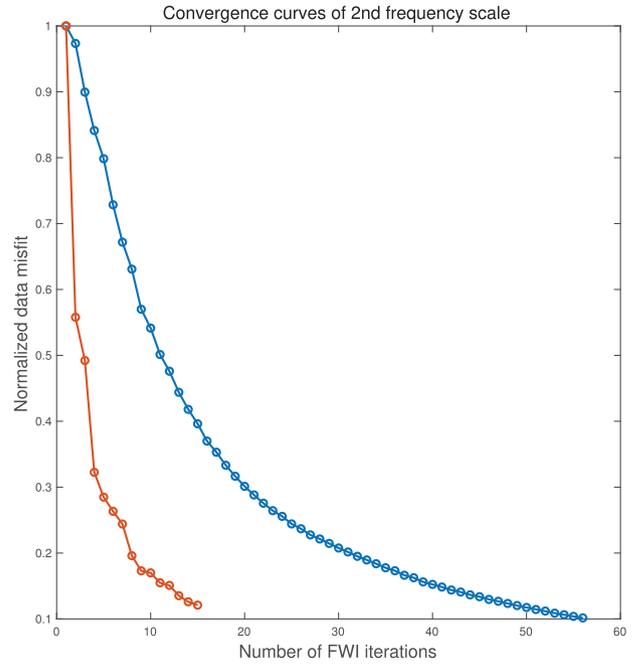


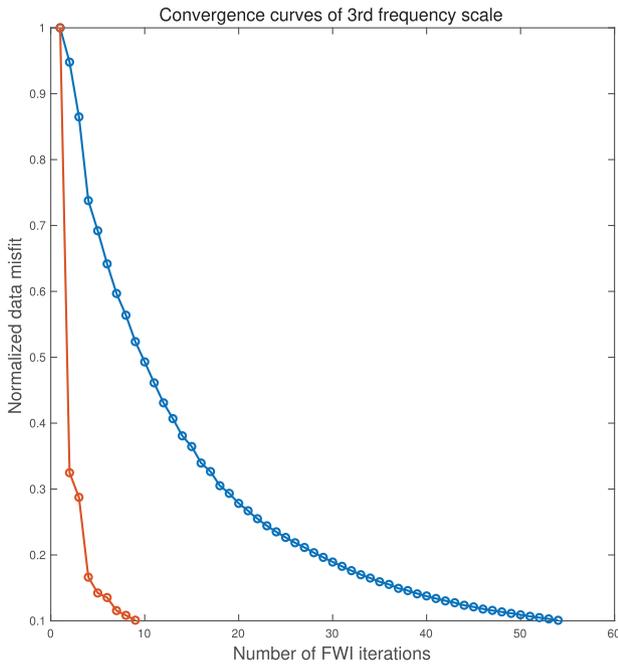
Figure 22. Data residuals comparison for shot at $x = 1500$ m. The data are plotted with heavy clipping (one fifth of the scale used in Figs 20 and 21). (a) Horizontal component data residual of elastic NLCG FWI. (b) Vertical component data residual of elastic NLCG FWI. (c) Horizontal component data residual of elastic Gauss–Newton FWI. (d) Vertical component data residual of elastic Gauss–Newton FWI.



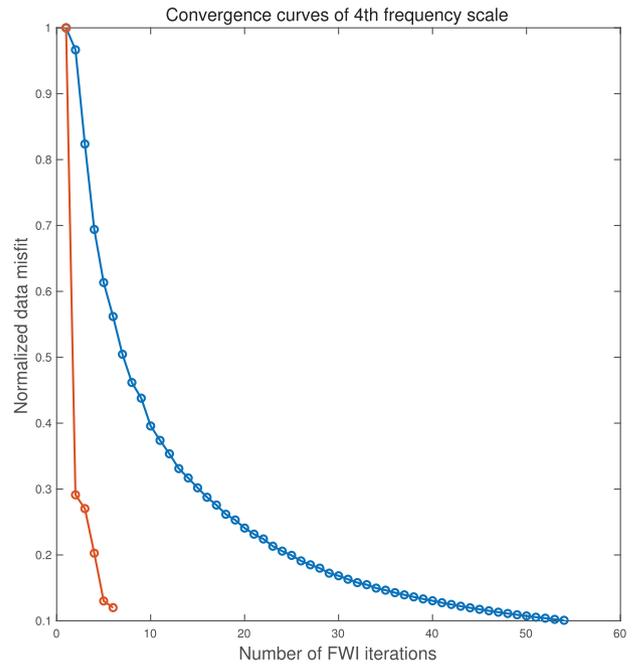
(a)



(b)



(c)



(d)

Figure 23. Normalized data misfit versus number of non-linear FWI iterations for elastic NLCG FWI and elastic Gauss–Newton FWI in frequency band (a) 2–3 Hz, (b) 2–5 Hz, (c) 2–10 Hz, (d) 2–30 Hz. Blue: data misfit curves for elastic NLCG FWI. Red: data misfit curves for elastic Gauss–Newton FWI.

3.0 km in the horizontal axis and 1.3 km depth with 301×131 gridpoints. There are 61 shots and 301 receivers deployed along the surface that simulates a fixed-spread survey geometry. The shot interval is 50 m, and receiver interval is 10.0 m. A 10 Hz central frequency Ricker wavelet is used to simulate an explosive source. The forward modelling is second-order accurate in time and 12th-order accurate in space ($\mathcal{O}(\Delta t^2, \Delta x^{12})$). The 0–2 Hz frequency content is removed from the data using a minimum phase Butterworth filter. We mute near-offset traces to avoid large amplitudes at receiver positions close to a source. The four frequency bands for the multiscale inversion are 2–3 Hz, 2–5 Hz,

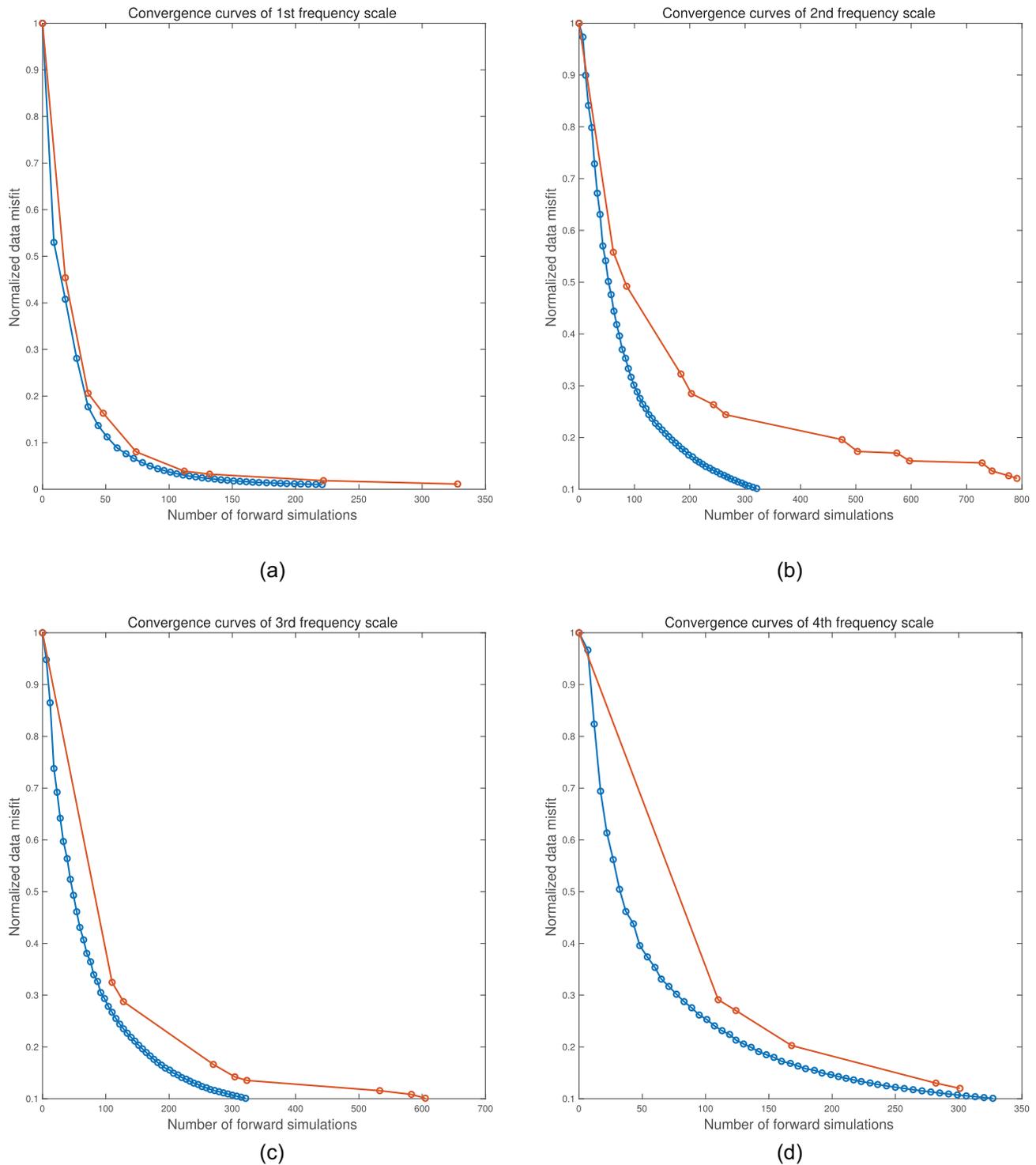


Figure 24. Normalized data misfit versus number of forward simulations for elastic NLCG FWI and elastic Gauss–Newton FWI in frequency band (a) 2–3 Hz, (b) 2–5 Hz, (c) 2–10 Hz, (d) 2–30 Hz. Blue: data misfit curves for elastic NLCG FWI. Red: data misfit curves for elastic Gauss–Newton FWI.

2–10 Hz and 2–30 Hz. Fig. 14 shows the starting P - and S -wave velocity models for the elastic NLCG FWI and the elastic Gauss–Newton FWI. The initial models are 1-D with linearly increasing P - and S -wave velocities.

For both optimization methods, we set the stopping criterion of the non-linear FWI iterative loop to $\epsilon = 0.01, 0.1, 0.1, 0.1$ for the four different frequency bands, respectively. The stopping criterion for the inner loop of the Gauss–Newton method is set to $\xi = 0.5$, and the maximum number of iterations allowed for the inner loop is 50. The parameters used in the inversion are summarized in Table 2.

The results of elastic NLCG FWI are shown in Figs 15(a) and 16(a). The results of elastic Gauss–Newton FWI are displayed in Figs 15(b) and 16(b). The results of the two optimization methods are similar. We compare the profiles of the inverted P -wave velocity models in Fig. 17,

and compare the profiles of the inverted S -wave velocity models in Fig. 18. The models inverted by the Gauss–Newton method appear to approximate the actual models more closely than the NLCG method. However, we recognize that these differences are minimal. In Fig. 19, we plot the model update direction of the two methods in the first non-linear FWI iteration in the first frequency band (2–3 Hz). The model update direction in the Gauss–Newton method has higher uniformity in its amplitude response than the NLCG method.

Next, we compare how well the inverted models by the two methods fit the observed data. Fig. 20 presents the data fitting of the elastic NLCG FWI. Fig. 21 presents the data fitting of the elastic Gauss–Newton FWI. The data prediction of both methods is generally good. To reveal more information, we plot the data residuals with heavy clipping in Fig. 22 (one fifth of the scale used in Figs 20 and 21). The elastic Gauss–Newton FWI provides a slightly better data prediction than the elastic NLCG FWI.

The convergence curves are plotted in terms of the normalized data misfit versus the number of the non-linear iterations. Fig. 23 compares the convergence curves of the two methods for four different frequency bands. The Gauss–Newton method has a faster convergence rate than the NLCG method. We compare the computational cost of the two methods by plotting curves of the normalized data misfit versus the number of forward simulations that were performed by each method. Fig. 24 compares the computational cost of the two methods for four different frequency bands. The total number of forward simulations performed by elastic NLCG FWI and our elastic Gauss–Newton FWI is 1190 and 2025, respectively. Under this configuration, the computational cost of elastic Gauss–Newton FWI is about 1.7 times of the cost of elastic NLCG FWI. For 2-D problems of the size explored by our paper, the differential cost is not high, and one can say that the Gauss–Newton inversion stands as a valid alternative for NLCG FWI

4 DISCUSSION

The Gauss–Newton method developed by us and presented in the paper is not supplying significant improvement in recovering the P - and S -wave velocities to the classical non-linear conjugate gradient method. Theoretically, different optimization methods should converge to the same point if a sufficiently large number of iterations is allowed, and we initially thought that convergence differences would compensate for the extra workload of the Gauss–Newton solver. The latter does not preclude the present research and numerical examples from being part of a body of research that expands our understanding of algorithms for FWI. A vast and rich literature is available describing Gauss–Newton and Newton’s optimization methods for FWI problems (Epanomeritakis *et al.* 2008; Heinkenschloss 2008; Metivier *et al.* 2013, 2014, 2017; Yang *et al.* 2018). To the best of our knowledge, however, our paper is the first presenting simultaneous inversion of P - and S -wave velocities using a matrix-free Gauss–Newton method in the time domain. In our opinion, the Gauss–Newton algorithm might not be required for recovering the background models when computational resources are limited. One can resort to less expensive optimization techniques such as NLCG. For recovering localized model anomalies, the matrix-free elastic Gauss–Newton FWI generates slightly better results than elastic NLCG FWI. The elastic Gauss–Newton FWI algorithm is more expensive than the elastic NLCG FWI. This is because the elastic Gauss–Newton FWI algorithm has an additional inner loop (CGLS) loop to estimate the search direction. There are several strategies to reduce the computational cost of elastic Gauss–Newton FWI. One can adopt data contraction strategies such as source encoding, plane-wave synthesis or stochastic source subsampling to reduce the computational cost (Vigh & Starr 2008a; Krebs *et al.* 2009; Schuster *et al.* 2011; van Leeuwen & Herrmann 2013; Matharu & Sacchi 2018). Alternatively, one could explore adopting immediate termination of the CGLS algorithm and explore if cost-benefit tradeoffs are more palatable for real large-scale applications. Or the elastic FWI could begin with non-linear conjugate gradient algorithm and switch over to Gauss–Newton method in the process based on the trade-off between convergence property gain and computational cost of the Gauss–Newton algorithm (Burstedde & Ghattas 2009). We also point out that in this paper, we do not address regularization strategies for FWI. We are aware that different regularization techniques may generate improvements under specific conditions (Asnaashari *et al.* 2013; Xue *et al.* 2017; Anagaw & Sacchi 2018b). The crosstalk between P - and S -wave velocities is not as severe as the crosstalk between density and velocities in the elastic full-waveform inversion. The conventional elastic NLCG FWI can achieve acceptable results in reconstructing the P - and S -wave velocities. The Gauss–Newton algorithm has the potential to reduce the crosstalk between the density and velocities in elastic FWI which provides a future research direction.

5 CONCLUSIONS

We present a time-domain matrix-free elastic Gauss–Newton FWI algorithm. We formulate the elastic Gauss–Newton FWI as an iterative elastic LSRTM problem. The proposed algorithm consists of two iterative loops: the outer Gauss–Newton iterations to update the model and the inner CGLS iterations to estimate the Gauss–Newton direction. The external iteration uses a parabolic fitting line search to determine the step size. The Gauss–Newton search direction is computed using the matrix-free CGLS algorithm. We point out that this step is equivalent to apply an elastic LSRTM to data residuals with the Fréchet derivative operator as an elastic Born forward modelling operator and the adjoint of Fréchet derivative operator as the elastic RTM operator. The CGLS method is preconditioned via the elastic pseudo-Hessian operator. Our algorithm is matrix-free as it only requires the forward Fréchet derivative and adjoint Fréchet derivative operators applied in implicit-form to vectors. The operators are applied on vectors efficiently via the adjoint-state method without having to store matrices in explicit form. The proposed algorithm is adopted to invert for P - and S -wave velocities simultaneously. Our research has led to the following observations. The proposed elastic Gauss–Newton FWI generates slightly better-inverted models than the elastic NLCG FWI method, which was used as a representative solver for mid and large size FWI problem. Despite the observed marginal improvements concluded from our simulations,

the time-domain Gauss–Newton FWI method described in this paper provides a strong formal connection to current research on seismic migration via LSRTM. Moreover, the framework described in the paper gives significant freedom to design Gauss–Newton FWI algorithms from existing elastic LSRTM codes that are emerging in the field of exploration seismology.

ACKNOWLEDGEMENTS

We thank the sponsors of the Signal Analysis and Imaging Group (SAIG) at the University of Alberta for financial support. This research is also supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) via a Discovery Grant to MDS. Compute Canada provided high-performance computing resources.

REFERENCES

- Akcelik, V., Biros, G. & Ghattas, O., 2002. Parallel multiscale Gauss–Newton–Krylov methods for inverse wave propagation, in *Proceedings of the Supercomputing, ACM/IEEE 2002 Conference*, pp. 41–41.
- Alemie, W. & Sacchi, M., 2016. Joint reparametrized time-lapse full-waveform inversion, in *Proceedings of the 86th Annual International Meeting, SEG, Expanded Abstracts*, pp. 1309–1314.
- Anagaw, A.Y. & Sacchi, M.D., 2018a. Model parametrization strategies for newton-based acoustic full waveform inversion, *J. appl. Geophys.*, **157**, 23–36.
- Anagaw, A.Y. & Sacchi, M.D., 2018b. Edge-preserving smoothing for simultaneous-source full-waveform inversion model updates in high-contrast velocity models, *Geophysics*, **83**(2), A33–A37.
- Anikiev, D., Kashtan, B. & Mulder, W.A., 2013. Decoupling of elastic parameters with iterative linearized inversion, in *Proceedings of the 83rd Annual International Meeting, SEG, Expanded Abstracts*, pp. 3185–3190.
- Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P. & Virieux, J., 2013. Regularized seismic full waveform inversion with prior model information, *Geophysics*, **78**(2), R25–R36.
- Bamberger, A., Chavent, G. & Lailly, P., 1979. About the stability of the inverse problem in 1-D wave equations - application to the interpretation of seismic profiles, *Appl. Math. Opt.*, **5**(1), 1–47.
- Bamberger, A., Chavent, G., Hemon, C. & Lailly, P., 1982. Inversion of normal incidence seismograms, *Geophysics*, **47**(5), 757–770.
- Baumstein, A., 2014. Extended subspace method for attenuation of crosstalk in multi-parameter full wavefield inversion, in *Proceedings of the 84th Annual International Meeting, SEG, Expanded Abstracts*, pp. 1121–1125.
- Björck, A., 1996. *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics.
- Borisov, D. & Singh, S.C., 2015. Three-dimensional elastic full waveform inversion in a marine environment using multicomponent ocean-bottom cables: a synthetic study, *Geophys. J. Int.*, **201**(3), 1215–1234.
- 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.
- Borzi, A. & Schulz, V., 2012. *Computational Optimization of Systems Governed by Partial Differential Equations*, Society for Industrial and Applied Mathematics.
- Brandsberg-Dahl, S., Chemingui, N., Valenciano, A., Ramos-Martinez, J. & Qiu, L., 2017. FWI for model updates in large-contrast media, *Leading Edge*, **36**(1), 81–87.
- Brossier, R., Operto, S. & Virieux, J., 2015. Velocity model building from seismic reflection data by full-waveform inversion, *Geophys. Prospect.*, **63**(2), 354–367.
- Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**(5), 1457–1473.
- Burstedde, C. & Ghattas, O., 2009. Algorithmic strategies for full waveform inversion: 1-D experiments, *Geophysics*, **74**(6), WCC37–WCC46.
- Chen, K. & Sacchi, M.D., 2017. Elastic least-squares reverse time migration via linearized elastic full-waveform inversion with pseudo-Hessian preconditioning, *Geophysics*, **82**(5), S341–S358.
- Chen, K. & Sacchi, M.D., 2018. Time-domain elastic Gauss–Newton full-waveform inversion: a matrix-free approach, in *Proceedings of the 88th Annual International Meeting, SEG, Expanded Abstracts*, pp. 1208–1212.
- Chen, K. & Sacchi, M.D., 2019. The importance of including density in elastic least-squares reverse time migration: multiparameter crosstalk and convergence, *Geophys. J. Int.*, **216**(1), 61–80.
- Chen, P., Zhao, L. & Jordan, T.H., 2007. Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. seism. Soc. Am.*, **97**(4), 1094–1120.
- Claerbout, J.F., 1985. *Imaging the Earth's Interior*, Blackwell Scientific Publications, Inc.
- Claerbout, J.F., 1992. *Earth Soundings Analysis: Processing Versus Inversion*, Vol. 6, Blackwell Scientific Publications.
- Duan, Y., Guitton, A. & Sava, P., 2017. Elastic least-squares reverse time migration, *Geophysics*, **82**(4), S315–S325.
- Dussaud, E., Symes, W.W., Williamson, P., Lemaistre, L., Singer, P., Denel, B. & Cherrett, A., 2008. Computational strategies for reverse-time migration, in *Proceedings of the 78th Annual International Meeting, SEG, Expanded Abstracts*, pp. 2267–2271.
- Eisenstat, S. & Walker, H., 1996. Choosing the forcing terms in an inexact newton method, *SIAM J. Sci. Comput.*, **17**(1), 16–32.
- Epanomeritakis, I., Akcelik, V., Ghattas, O. & Bielak, J., 2008. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion, *Inverse Problems*, **24**(3), 034015.
- Feng, Z. & Schuster, G.T., 2017. Elastic least-squares reverse time migration, *Geophysics*, **82**(2), S143–S157.
- Fichtner, A. & Trampert, J., 2011. Hessian kernels of seismic data functionals based upon adjoint techniques, *Geophys. J. Int.*, **185**(2), 775–798.
- Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology: I. theory, *Phys. Earth planet. Inter.*, **157**(1–2), 86–104.
- Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.-P., 2009. Full seismic waveform tomography for upper-mantle structure in the australasian region using adjoint methods, *Geophys. J. Int.*, **179**(3), 1703–1725.
- Forgues, E. & Lambare, G., 1997. Parameterization study for acoustic and elastic ray + Born inversion, *J. Seism. Explor.*, **6**, 253–277.
- Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional non-linear inversion of seismic waveforms: Numerical results, *Geophysics*, **51**(7), 1387–1403.
- Gratton, S., Lawless, A. & Nichols, N., 2007. Approximate Gauss–Newton methods for non-linear least squares problems, *SIAM J. Opt.*, **18**(1), 106–132.
- Graves, R.W., 1996. Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences, *Bull. seism. Soc. Am.*, **86**(4), 1091.
- Gu, B., Li, Z., Yang, P., Xu, W. & Han, J., 2017. Elastic least-squares reverse time migration with hybrid 11/12 misfit function, *Geophysics*, **82**(3), S271–S291.
- Guasch, L., Warner, M., Nangoo, T., Morgan, J., Umpleby, A., Stekl, I. & Shah, N., 2012. Elastic 3D full-waveform inversion, in *Proceedings of the 82nd Annual International Meeting, SEG, Expanded Abstracts*, pp. 1–5.
- Heinkenschloss, M., 2008. Numerical solution of implicitly constrained optimization problems, CAAM Technical Report TR08-05.
- Hestenes, M.R. & Stiefel, E., 1952. Methods of conjugate gradients for solving linear systems, *J. Res. Natl. Bureau Stand.*, **49**(6), 409–436.
- Kamei, R. & Lumley, D., 2017. Full waveform inversion of repeating seismic events to estimate time-lapse velocity changes, *Geophys. J. Int.*, **209**(2), 1239–1264.

- Kohn, D., De Nil, D., Kurzmann, A., Przebindowska, A. & Bohlen, T., 2012. On the influence of model parametrization in elastic full waveform tomography, *Geophys. J. Int.*, **191**(1), 325–345.
- Kolb, P., Collino, F. & Lailly, P., 1986. Pre-stack inversion of a 1-D medium, *Proc. IEEE*, **74**(3), 498–508.
- Komatitsch, D. & Martin, R., 2007. An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation, *Geophysics*, **72**(5), SM155–SM167.
- Krebs, J.R., Anderson, J.E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A. & Lacasse, M.-D., 2009. Fast full-wavefield seismic inversion using encoded sources, *Geophysics*, **74**(6), WCC177–WCC188.
- Li, X., Esser, E. & Herrmann, F.J., 2016. Modified Gauss–Newton full-waveform inversion explained—why sparsity-promoting updates do matter, *Geophysics*, **81**(3), R125–R138.
- Lions, J.L., 1971. *Optimal Control of Systems Governed by Partial Differential Equations*, Springer Berlin Heidelberg.
- Liu, Y. & Sen, M.K., 2009. An implicit staggered-grid finite-difference method for seismic modelling, *Geophys. J. Int.*, **179**(1), 459–474.
- Lu, R., 2016. Revealing overburden and reservoir complexity with high-resolution FWI in *Proceedings of the 86th Annual International Meeting, SEG*, Expanded Abstracts, pp. 1242–1246.
- Luo, Y., Ma, Y., Wu, Y., Liu, H. & Cao, L., 2016. Full-traveltime inversion, *Geophysics*, **81**(5), R261–R274.
- Mao, J., Sheng, J., Hart, M. & Kim, T., 2016. High-resolution model building with multistage full-waveform inversion for narrow-azimuth acquisition data, *Leading Edge*, **35**(12), 1031–1036.
- Martin, G.S., Wiley, R. & Marfurt, K.J., 2006. Marmousi2: An elastic upgrade for Marmousi, *Leading Edge*, **25**(2), 156–166.
- Matharu, G. & Sacchi, M.D., 2018. Source encoding in multiparameter full waveform inversion, *Geophys. J. Int.*, **214**(2), 792–810.
- Metivier, L., Brossier, R., Virieux, J. & Operto, S., 2013. Full waveform inversion and the truncated Newton method, *SIAM J. Sci. Comput.*, **35**(2), B401–B437.
- Metivier, L., Bretteau, F., Brossier, R., Operto, S. & Virieux, J., 2014. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures, *Geophys. Prospect.*, **62**(6), 1353–1375.
- Metivier, L., Brossier, R., Operto, S. & Virieux, J., 2017. Full waveform inversion and the truncated Newton method, *SIAM Rev.*, **59**(1), 153–195.
- Mora, P., 1987a. Non-linear two-dimensional elastic inversion of multioffset seismic data, *Geophysics*, **52**(9), 1211–1228.
- Mora, P., 1987b. Elastic wavefield inversion, *PhD thesis*, Stanford University.
- Mora, P., 1988. Elastic wave-field inversion of reflection and transmission data, *Geophysics*, **53**(6), 750–759.
- Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization*, Springer.
- Oh, J.-W., Kalita, M. & Alkhalifah, T., 2018. 3D elastic full-waveform inversion using P-wave excitation amplitude: application to ocean bottom cable field data, *Geophysics*, **83**(2), R129–R140.
- Paige, C.C. & Saunders, M.A., 1982. LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw.*, **8**(1), 43–71.
- Pan, W., Innanen, K.A., Margrave, G.F., Fehler, M.C., Fang, X. & Li, J., 2016. Estimation of elastic constants for HTI media using Gauss–Newton and full-Newton multiparameter full-waveform inversion, *Geophysics*, **81**(5), R275–R291.
- 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.
- Plessix, R.-E. & Perkins, C., 2010. Thematic set: Full waveform inversion of a deep water ocean bottom seismometer dataset, *First Break*, **28**(4), 71–78.
- Pratt, R.G., Shin, C. & Hick, G.J., 1998. Gauss–Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.*, **133**(2), 341–362.
- Raknes, E.B. & Arntsen, B., 2014. Time-lapse full-waveform inversion of limited-offset seismic data using a local migration regularization, *Geophysics*, **79**(3), WA117–WA128.
- Ren, Z., Liu, Y. & Sen, M.K., 2017. Least-squares reverse time migration in elastic media, *Geophys. J. Int.*, **208**(2), 1103–1125.
- Routh, P., Palacharla, G., Chikichev, I. & Lazaratos, S., 2012. Full wavefield inversion of time-lapse data for improved imaging and reservoir characterization, in *Proceedings of the 82nd Annual International Meeting, SEG*, Expanded Abstracts, pp. 1–6.
- Santosa, F. & Symes, W.W., 1988. Computation of the Hessian for least-squares solutions of inverse problems of reflection seismology, *Inverse Problems*, **4**(1), 211.
- Schuster, G.T., Wang, X., Huang, Y., Dai, W. & Boonyasiriwat, C., 2011. Theory of multisource crosstalk reduction by phase-encoded statics, *Geophys. J. Int.*, **184**(3), 1289–1303.
- Sears, T., Singh, S. & Barton, P., 2008. Elastic full waveform inversion of multi-component OBC seismic data, *Geophys. Prospect.*, **56**(6), 843–862.
- Sheen, D.-H., Tuncay, K., Baag, C.-E. & Ortoleva, P.J., 2006. Time domain Gauss–Newton seismic waveform inversion in elastic media, *Geophys. J. Int.*, **167**(3), 1373–1384.
- Shen, X., *et al.*, 2018. High-resolution full-waveform inversion for structural imaging in exploration, in *Proceedings of the 88th Annual International Meeting, SEG*, Expanded Abstracts, pp. 1098–1102.
- 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.
- Sirgue, L., Barkved, O., Dellinger, J., Etgen, J., Albertin, U. & Kommedal, J., 2010. Thematic set: full waveform inversion: the next leap forward in imaging at Valhall, *First Break*, **28**(4), 65–70.
- Sun, M., Dong, L., Yang, J., Huang, C. & Liu, Y., 2018. Elastic least-squares reverse time migration with density variations, *Geophysics*, **83**(6), S533–S547.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the Southern California crust, *Science*, **325**(5943), 988–992.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the Southern California crust based on spectral-element and adjoint methods, *Geophys. J. Int.*, **180**(1), 433–462.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.
- Tarantola, A., 1986. A strategy for non-linear elastic inversion of seismic reflection data, *Geophysics*, **51**(10), 1893–1903.
- Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *PAGEOPH*, **128**, 365–399.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics.
- Tarantola, A. & Valette, B., 1982. Generalized non-linear inverse problems solved using the least squares criterion, *Rev. Geophys.*, **20**(2), 219–232.
- Trinh, P.-T., Romain, Brossier, Metivier, L., Tavard, L. & Virieux, J., 2018. Efficient time-domain 3D elastic and visco-elastic FWI using a spectral-element method on flexible Cartesian-based mesh, *Geophysics*, **0**(ja), 1–91.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**(1), 195–216.
- van Leeuwen, T. & Herrmann, F.J., 2013. Fast waveform inversion without source-encoding, *Geophys. Prospect.*, **61**, 10–19.
- Vigh, D. & Starr, E.W., 2008a. 3D prestack plane-wave, full-waveform inversion, *Geophysics*, **73**(5), VE135–VE144.
- Vigh, D. & Starr, E.W., 2008b. Comparisons for waveform inversion, time domain or frequency domain?, in *Proceedings of the 78th Annual International Meeting, SEG*, Expanded Abstracts, pp. 1890–1894.
- Vigh, D., Starr, E.W. & Kapoor, J., 2009. Developing earth models with full waveform inversion, *Leading Edge*, **28**(4), 432–435.
- 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.
- Virieux, J., 1986. P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method, *Geophysics*, **51**(4), 889–901.

- Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
- Wang, C., Yingst, D., Farmer, P. & Leveille, J., 2016. Full-waveform inversion with the reconstructed wavefield method, in *Proceedings of the 86th Annual International Meeting*, , SEG, Expanded Abstracts, pp. 1237–1241.
- Witte, P., Louboutin, M., Lensink, K., Lange, M., Kukreja, N., Luporini, F., Gorman, G. & Herrmann, F.J., 2018. Full-waveform inversion, part 3: optimization, *Leading Edge*, **37**(2), 142–145.
- Xu, L., Stanton, A. & Sacchi, M., 2016. Elastic least-squares reverse time migration, in *Proceedings of the 86th Annual International Meeting*, SEG, Expanded Abstracts, pp. 2289–2293.
- Xu, S., Wang, D., Chen, F., Lambare, G. & Zhang, Y., 2012. Inversion on reflected seismic wave, in *Proceedings of the 82nd Annual International Meeting*, SEG, Expanded Abstracts, pp.1–7.
- Xue, Z., Zhu, H. & Fomel, S., 2017. Full-waveform inversion using seislet regularization, *Geophysics*, **82**(5), A43–A49.
- Yang, D., Liu, F., Morton, S., Malcolm, A. & Fehler, M., 2016. Time-lapse full-waveform inversion with ocean-bottom-cable data: application on Valhall field, *Geophysics*, **81**(4), R225–R235.
- Yang, P., Brossier, R., Metivier, L., Virieux, J. & Zhou, W., 2018. A time-domain preconditioned truncated newton approach to visco-acoustic multiparameter full waveform inversion, *SIAM J. Sci. Comput.*, **40**(4), B1101–B1130.

A1 DERIVATION OF THE FORWARD AND ADJOINT FRÉCHET DERIVATIVE OPERATORS

A1.1 Derivation of the Fréchet derivative operator L

In this section, we drop the source index i in the Fréchet derivative operator to make the expressions more concise. The Fréchet derivative of the objective function of the FWI problem is the Born approximation modelling operator. The latter maps model perturbation to data perturbation

$$\delta \mathbf{d} = \mathbf{L} \delta \mathbf{m} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}} \delta \mathbf{m} = \mathbf{R} \frac{\partial \mathbf{u}}{\partial \mathbf{m}} \delta \mathbf{m}. \quad (\text{A1})$$

The partial derivative $\partial \mathbf{u} / \partial \mathbf{m}$ is prohibitively expensive to compute explicitly. It can be replaced using the implicit function theorem. Differentiating the wave eq. (5) with respect to the vector of model parameters \mathbf{m} yields

$$\frac{\partial \mathbf{S}}{\partial \mathbf{m}} \mathbf{u} + \mathbf{S} \frac{\partial \mathbf{u}}{\partial \mathbf{m}} = \mathbf{0}. \quad (\text{A2})$$

The right side of the equation becomes zero because the source does not depend on \mathbf{m} . Inserting the expression of the partial derivative $\partial \mathbf{u} / \partial \mathbf{m}$ into eq. (A1) leads to the Born approximation of a general wave eq. (5)

$$\delta \mathbf{d} = \mathbf{L} \delta \mathbf{m} = -\mathbf{R} \mathbf{S}^{-1} \left(\frac{\partial \mathbf{S}}{\partial \mathbf{m}} \mathbf{u} \right) \delta \mathbf{m} = -\mathbf{R} \mathbf{S}^{-1} \left[\frac{\partial \mathbf{S}}{\partial \mathbf{m}} (\mathbf{S}^{-1} \mathbf{f}) \right] \delta \mathbf{m}, \quad (\text{A3})$$

where \mathbf{u} represents the incident wavefield. Moreover, the solution of the wave equation $\mathbf{u} = \mathbf{S}^{-1} \mathbf{f}$ was applied.

We now show the Born approximation for the isotropic elastic wave equation system (4). Let us consider the source-side wavefield \mathbf{u} defined by the elastic wave equation

$$\begin{pmatrix} \rho \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \frac{\partial}{\partial t} \mathbf{u} - \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C} \mathbf{D}^T & \mathbf{0} \end{pmatrix} \mathbf{u} = \mathbf{f}. \quad (\text{A4})$$

A perturbation of the model parameters

$$\begin{aligned} \rho &\rightarrow \rho + \delta \rho, \\ \mathbf{C} &\rightarrow \mathbf{C} + \delta \mathbf{C}, \end{aligned} \quad (\text{A5})$$

leads to a perturbation of the wavefield

$$\mathbf{u} \rightarrow \mathbf{u} + \delta \mathbf{u}. \quad (\text{A6})$$

We substitute the perturbed model parameters and wavefields into eq. (A4). After canceling second order terms involving small perturbations, one ends up with the following partial differential equations

$$\begin{pmatrix} \rho \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \frac{\partial}{\partial t} \delta \mathbf{u} - \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C} \mathbf{D}^T & \mathbf{0} \end{pmatrix} \delta \mathbf{u} = \begin{pmatrix} -\delta \rho \frac{\partial}{\partial t} \mathbf{v} \\ \delta \mathbf{C} \mathbf{D}^T \mathbf{v} \end{pmatrix}, \quad (\text{A7})$$

where

$$\delta \mathbf{u} = \begin{pmatrix} \delta \mathbf{v} \\ \delta \boldsymbol{\sigma} \end{pmatrix}, \delta \mathbf{v} = \begin{pmatrix} \delta v_x \\ \delta v_y \\ \delta v_z \end{pmatrix}, \delta \boldsymbol{\sigma} = \begin{pmatrix} \delta \sigma_{xx} \\ \delta \sigma_{yy} \\ \delta \sigma_{zz} \\ \delta \sigma_{yz} \\ \delta \sigma_{xz} \\ \delta \sigma_{xy} \end{pmatrix}, \quad (\text{A8})$$

$$\delta \mathbf{C} = \begin{pmatrix} \delta\lambda + 2\delta\mu & \delta\lambda & \delta\lambda & 0 & 0 & 0 \\ \delta\lambda & \delta\lambda + 2\delta\mu & \delta\lambda & 0 & 0 & 0 \\ \delta\lambda & \delta\lambda & \delta\lambda + 2\delta\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta\mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \delta\mu \end{pmatrix},$$

the vector $\delta \mathbf{u}$ is the scattered wavefield caused by model perturbations $(\delta\rho, \delta\mathbf{C})$, and \mathbf{v} is the incident particle velocity field in the unperturbed model (ρ, \mathbf{C}) . eq. (A7) represents the first-order Born approximation in a heterogeneous, isotropic elastic media. The Born approximation equation can be solved using the same numerical solver that is used for solving the wave equation (eq. 4). The scattered data are obtained by sampling the scattered wavefield at the receiver positions

$$\delta \mathbf{d} = \mathbf{R} \delta \mathbf{u}. \quad (\text{A9})$$

For the elastic FWI problem, the parameterization in compressional and shear wave speeds is better than a parameterization in Lamé parameters (Tarantola 1986; Kohn *et al.* 2012) and this is the reason for selecting P- and S-wave velocities to invert for. By adopting the chain rule in the Fréchet derivative, one can transform perturbations accordingly to

$$\begin{pmatrix} \delta\rho \\ \delta\lambda \\ \delta\mu \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ V_p^2 - 2V_s^2 & 2\rho V_p & -4\rho V_s \\ V_s^2 & 0 & 2\rho V_s \end{pmatrix} \begin{pmatrix} \delta\varrho \\ \delta V_p \\ \delta V_s \end{pmatrix}, \quad (\text{A10})$$

where $\mathbf{m} = (\rho, V_p, V_s)^T$ represents the background model and $\delta \mathbf{m} = (\delta\varrho, \delta V_p, \delta V_s)^T$ the model perturbation. eqs (A4), (A7), (A9) and (A10) together define the Born approximation for elastic wave equation system (general form as eq. (A1)).

A1.2 Derivation of the adjoint of Fréchet derivative operator \mathbf{L}^\dagger

The adjoint of the Fréchet derivative of the elastic FWI objective function is the elastic RTM operator. The latter is the adjoint of the elastic Born modelling operator that maps data perturbation to model perturbation. If we take the adjoint of the eq. (A1), we can obtain

$$\begin{aligned} \delta \mathbf{m}^* &= \mathbf{L}^\dagger \delta \mathbf{d} = \left(\frac{\partial \mathbf{d}}{\partial \mathbf{m}} \right)^\dagger \delta \mathbf{d} = - \left(\frac{\partial \mathbf{S}}{\partial \mathbf{m}} \mathbf{u} \right)^\dagger (\mathbf{S}^{-1})^\dagger \mathbf{R}^\dagger \delta \mathbf{d} \\ &= - \left(\frac{\partial \mathbf{S}}{\partial \mathbf{m}} \mathbf{u} \right)^\dagger (\mathbf{S}^\dagger)^{-1} \mathbf{R}^\dagger \delta \mathbf{d} = - \left[\frac{\partial \mathbf{S}}{\partial \mathbf{m}} (\mathbf{S}^{-1} \mathbf{f}) \right]^\dagger (\mathbf{S}^\dagger)^{-1} \mathbf{R}^\dagger \delta \mathbf{d}, \end{aligned} \quad (\text{A11})$$

where $\delta \mathbf{m}^*$ is the adjoint model perturbation ($\delta \mathbf{m}^* = (\delta\varrho^*, \delta V_p^*, \delta V_s^*)^T$), \mathbf{L}^\dagger is the elastic RTM operator, and \mathbf{R}^\dagger is the adjoint of sampling operator. The elastic RTM operator (adjoint of Fréchet derivative operator) also contains two wave propagation operators: one forward (\mathbf{S}^{-1}) and one backward ($[\mathbf{S}^\dagger]^{-1}$) propagators. We introduce the adjoint-state variable $\mathbf{p} = (\mathbf{S}^\dagger)^{-1} \mathbf{R}^\dagger \delta \mathbf{d}$. The latter satisfies the adjoint-state equation corresponding to the general state eq. (5)

$$\mathbf{S}^\dagger \mathbf{p} = \mathbf{R}^\dagger \delta \mathbf{d}, \quad (\text{A12})$$

where \mathbf{p} is the adjoint-state variable of the state variable \mathbf{u} , \mathbf{S}^\dagger is the adjoint wave equation operator and $\mathbf{R}^\dagger \delta \mathbf{d}$ is the adjoint source. The estimated model perturbations or images can be simplified

$$\delta \mathbf{m}^* = - \left(\frac{\partial \mathbf{S}}{\partial \mathbf{m}} \mathbf{u} \right)^\dagger \mathbf{p}, \quad (\text{A13})$$

or rewritten the implicit inner product over time explicitly

$$\delta \mathbf{m}(\mathbf{x})^* = - \int \left(\frac{\partial \mathbf{S}(\mathbf{x}, t)}{\partial \mathbf{m}(\mathbf{x})} \mathbf{u}(\mathbf{x}, t) \right)^\dagger \mathbf{p}(\mathbf{x}, t) dt = - \int \mathbf{p}(\mathbf{x}, t)^T \left(\frac{\partial \mathbf{S}(\mathbf{x}, t)}{\partial \mathbf{m}(\mathbf{x})} \mathbf{u}(\mathbf{x}, t) \right) dt, \quad (\text{A14})$$

where $\mathbf{p}(\mathbf{x}, t)^\dagger = \mathbf{p}(\mathbf{x}, t)^T$. The model perturbation kernels obtained using the mathematical adjoint-state method are similar to Claerbout's physical zeros-lag cross-correlation imaging condition (Claerbout 1985). The only difference is that expression A14 has an additional scattering operator (the operator $\partial \mathbf{S}(\mathbf{x}, t) / \partial \mathbf{m}(\mathbf{x})$) applied on the wavefield. After applying the adjoint wave equation operator \mathbf{S}^\dagger to the

first-order velocity-stress elastic wave equation system 4, we arrive the following adjoint-state wave equation

$$\begin{pmatrix} \rho \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}^\dagger \left(\frac{\partial}{\partial t} \right)^\dagger \mathbf{p} - \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C} \mathbf{D}^T & \mathbf{0} \end{pmatrix}^\dagger \mathbf{p} = \mathbf{R}^\dagger \delta \mathbf{d}, \tag{A15}$$

where

$$\mathbf{p} = \begin{pmatrix} \mathbf{v} \\ \boldsymbol{\tau} \end{pmatrix}, \mathbf{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}, \boldsymbol{\tau} = \begin{pmatrix} \tau_{xx} \\ \tau_{yy} \\ \tau_{zz} \\ \tau_{yz} \\ \tau_{xz} \\ \tau_{xy} \end{pmatrix}, \delta \mathbf{d} = \begin{pmatrix} \delta d_{v_x} \\ \delta d_{v_y} \\ \delta d_{v_z} \\ \delta d_{\sigma_{xx}} \\ \delta d_{\sigma_{yy}} \\ \delta d_{\sigma_{zz}} \\ \delta d_{\sigma_{yz}} \\ \delta d_{\sigma_{xz}} \\ \delta d_{\sigma_{xy}} \end{pmatrix}, \tag{A16}$$

where \mathbf{p} is the adjoint-state wavefield with v the adjoint variable associated to particle velocity and τ as the adjoint variable associate to the stress field. The data vector $\delta \mathbf{d}$ acts as the adjoint source of the adjoint-state equation. The final condition $\mathbf{p}|_{t=T} = \mathbf{0}$ is imposed, and a free surface boundary condition on the top boundary and absorbing boundary conditions on the other boundaries of the model are imposed on the adjoint-state wave equation. The differential operator is anti-self-adjoint: $\left(\frac{\partial}{\partial x} \right)^\dagger = - \left(\frac{\partial}{\partial x} \right)$. Consequently, one can write the expression $\mathbf{D}^\dagger = -\mathbf{D}^T$. The adjoint-state eq. (A15) changes to

$$\begin{pmatrix} \rho \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \left(-\frac{\partial}{\partial t} \right) \mathbf{p} + \begin{pmatrix} \mathbf{0} & \mathbf{D} \mathbf{C} \\ \mathbf{D}^T & \mathbf{0} \end{pmatrix} \mathbf{p} = \mathbf{R}^\dagger \delta \mathbf{d}. \tag{A17}$$

The wave equation operator in the adjoint-state equation is slightly different to that of the state eq. (4). The adjoint-state equation can be redefined into a form that resembles the state equation using a transformation of variables

$$\tilde{\mathbf{p}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} \mathbf{p},$$

where $\tilde{\mathbf{p}} = (\tilde{\mathbf{v}}, \tilde{\boldsymbol{\tau}})^T$ is the transformed adjoint-state variable, $\tilde{\mathbf{v}} = (\tilde{v}_x, \tilde{v}_y, \tilde{v}_z)^T$ and $\tilde{\boldsymbol{\tau}} = (\tilde{\tau}_{xx}, \tilde{\tau}_{yy}, \tilde{\tau}_{zz}, \tilde{\tau}_{yz}, \tilde{\tau}_{xz}, \tilde{\tau}_{xy})^T$. Multiplying both sides of the adjoint-state eq. (A17) by the transformation matrix, the adjoint-state equation becomes

$$\begin{pmatrix} \rho \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \left(-\frac{\partial}{\partial t} \right) \tilde{\mathbf{p}} + \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C} \mathbf{D}^T & \mathbf{0} \end{pmatrix} \tilde{\mathbf{p}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} \mathbf{R}^\dagger \delta \mathbf{d}, \tag{A18}$$

where the adjoint-state equation now has the same structure as the state eq. (4). The same numerical solver adopted in solving the elastic wave equation system 4 can be reused to solve the adjoint wave equation (eq. A18). The difference is that time evolution is in reverse-time direction in the adjoint wave equation. After computing the transformed adjoint-state variable $\tilde{\mathbf{p}}$, the original adjoint-state variable \mathbf{p} can be retrieved by the inverse transformation

$$\mathbf{p} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{-1} \end{pmatrix} \tilde{\mathbf{p}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{-1} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{v}} \\ \tilde{\boldsymbol{\tau}} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{v}} \\ \mathbf{C}^{-1} \tilde{\boldsymbol{\tau}} \end{pmatrix}, \tag{A19}$$

with

$$\mathbf{C}^{-1} = \begin{pmatrix} \frac{\lambda + \mu}{3\lambda\mu + 2\mu^2} & -\frac{\lambda}{6\lambda\mu + 4\mu^2} & -\frac{\lambda}{6\lambda\mu + 4\mu^2} & 0 & 0 & 0 \\ -\frac{6\lambda\mu + 4\mu^2}{\lambda} & \frac{\lambda}{3\lambda\mu + 2\mu^2} & -\frac{6\lambda\mu + 4\mu^2}{\lambda + \mu} & 0 & 0 & 0 \\ -\frac{6\lambda\mu + 4\mu^2}{\lambda} & -\frac{6\lambda\mu + 4\mu^2}{\lambda} & \frac{6\lambda\mu + 4\mu^2}{3\lambda\mu + 2\mu^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\mu} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\mu} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\mu} \end{pmatrix}. \tag{A20}$$

Using the elastic wave equation (eq. 4), the adjoint wave equation (eq. A18) and the model perturbation equation (eq. A14), the adjoint model perturbations $(\delta \mathbf{m}^* = (\delta \rho^*, \delta \lambda^*, \delta \mu^*)^T)$ can be expressed as

$$\begin{aligned}
 \delta \rho^* &= - \int \dot{\mathbf{v}} \cdot \mathbf{v} dt = - \int \dot{\mathbf{v}} \cdot \tilde{\mathbf{v}} dt \\
 &= - \int (\dot{v}_x \tilde{v}_x + \dot{v}_y \tilde{v}_y + \dot{v}_z \tilde{v}_z) dt \\
 \delta \lambda^* &= \int \left(\frac{\partial \mathbf{C}}{\partial \lambda} \mathbf{D}^T \mathbf{v} \right) \cdot \boldsymbol{\tau} dt = \int \left(\frac{\partial \mathbf{C}}{\partial \lambda} \mathbf{C}^{-1} \dot{\boldsymbol{\sigma}} \right) \cdot (\mathbf{C}^{-1} \tilde{\boldsymbol{\tau}}) dt \\
 &= \int \frac{(\dot{\sigma}_{xx} + \dot{\sigma}_{yy} + \dot{\sigma}_{zz})(\tilde{\tau}_{xx} + \tilde{\tau}_{yy} + \tilde{\tau}_{zz})}{(3\lambda + 2\mu)^2} dt \\
 \delta \mu^* &= \int \left(\frac{\partial \mathbf{C}}{\partial \mu} \mathbf{D}^T \mathbf{v} \right) \cdot \boldsymbol{\tau} dt = \int \left(\frac{\partial \mathbf{C}}{\partial \mu} \mathbf{C}^{-1} \dot{\boldsymbol{\sigma}} \right) \cdot (\mathbf{C}^{-1} \tilde{\boldsymbol{\tau}}) dt \\
 &= \int \left[-\frac{3\lambda^2 + 4\lambda\mu}{2\mu^2(3\lambda + 2\mu)^2} (\dot{\sigma}_{xx} + \dot{\sigma}_{yy} + \dot{\sigma}_{zz})(\tilde{\tau}_{xx} + \tilde{\tau}_{yy} + \tilde{\tau}_{zz}) \right. \\
 &\quad \left. + \frac{1}{2\mu^2} (\dot{\sigma}_{xx} \tilde{\tau}_{xx} + \dot{\sigma}_{yy} \tilde{\tau}_{yy} + \dot{\sigma}_{zz} \tilde{\tau}_{zz}) + \frac{1}{\mu^2} (\dot{\sigma}_{yz} \tilde{\tau}_{yz} + \dot{\sigma}_{xz} \tilde{\tau}_{xz} + \dot{\sigma}_{xy} \tilde{\tau}_{xy}) \right] dt, \tag{A21}
 \end{aligned}$$

where the over-dot means time derivative.

The wave speeds are easier to resolve than the Lamé parameters in elastic FWI (Tarantola 1986; Forgues & Lambare 1997). The chain rule can connect different parameter perturbations in elastic FWI

$$\begin{pmatrix} \delta \rho^* \\ \delta V_p^* \\ \delta V_s^* \end{pmatrix} = \begin{pmatrix} 1 & V_p^2 - 2V_s^2 & V_s^2 \\ 0 & 2\rho V_p & 0 \\ 0 & -4\rho V_s & 2\rho V_s \end{pmatrix} \begin{pmatrix} \delta \rho^* \\ \delta \lambda^* \\ \delta \mu^* \end{pmatrix}. \tag{A22}$$

eqs (A18), (A21) and (A22) together define the adjoint Fréchet operator (elastic RTM operator).