Multi-parameter acoustic full-waveform inversion: a comparison of different parameterizations and optimization methods

Wenyong Pan, Kris Innanen, Yu Geng

ABSTRACT

Full-waveform inversion methods allow to provide high-resolution estimates of subsurface elastic properties, which are very important for reservoir characterization. However, multi-parameter FWI suffers from parameter crosstalk artifacts arising from the inherent ambiguities (or coupling effects) among different physical parameters, which significantly increase the non-linearity of the inverse problems. For multi-parameter acoustic FWI, density is difficult to be inverted, which maybe caused by the strong parameter trade-off from velocity. Different parameterizations have different resolving abilities. An appropriate parameterization in multi-parameter FWI can help avoid parameter crosstalk and reconstruct the model parameters efficiently. Scattering patterns due to different physical parameters have been employed to study the resolving abilities of different parameterizations. In this research, we will illustrate the scattering patterns for different parameterizations in acoustic media. It has been proved that the second-order derivative (namely Hessian operator) is capable to suppress the parameter crosstalk artifacts. The Hessian-free optimization methods are generally employed for avoiding construct the multi-parameter Hessian explicitly. In this research, we compare the inversion efficiencies of different parameterizations with different optimization methods for multi-parameter acoustic FWI.

INTRODUCTION

Full-waveform inversion (FWI) methods are promising techniques for providing high-resolution estimates of subsurface elastic and anisotropic properties (Lailly, 1983; Tarantola, 1984; Pratt et al., 1998; Virieux and Operto, 2009). In recent decades, geophysicists have devoted significant efforts for applying these methods to reconstruct subsurface models in oil and gas exploration. However, even the mono-parameter FWI, which only tries to invert P-wave velocity, is difficult to be applied successfully because of cycle-skipping problem arising from lack of low frequencies and inaccurate initial models. Researchers have developed various strategies including joint travel time and waveform inversion methods (Ma and Hale, 2012), envelope misfit function, adaptive matching filtering and extrapolated frequencies (Li and Demanet, 2016; Pan et al., 2016c).

FWI methods should also be employed for recovering the subsurface elastic properties, which are essential for reservoir characterization. The parameter crosstalk artifacts involved by the inherent ambiguities (or coupling effects) among different physical parameters significantly increase the non-linearity of the inverse problems (Operto et al., 2013; Innanen, 2014a). Density is important fluid reservoir prediction. However, density is very difficult to be reconstructed due to its insensitivity to travel time and parameter trade-offs from P-wave and S-wave velocities. In elastic FWI, the strong mapping from S-wave velocity perturbation to density make density difficult to be recovered. In acoustic FWI when considering variable density, the strong parameter crosstalk between P-wave velocity and

density make it difficult to invert density.

In multi-parameter FWI, different parameterizations have different resolving abilities. An appropriate parameterization should have high parameter resolution, which is essential for effective inversion. Parameter resolution issue based on scattering patterns have been widely studied for selecting the optimal parameterization for avoiding parameter trade-off. The scattering patterns represent the analytic solutions of the Fréchet derivative wavefields due to different physical parameters. The overlapping of the scattering patterns due to different physical parameters indicates the appearance of the parameter crosstalk. In this research, we will examine the efficiencies of different parameterizations for multi-parameter acoustic FWI.

The second-order partial derivative of the misfit function (namely the Hessian operator) carries crucial information in the reconstruction process (Santosa and Symes, 1988; Fichtner and Trampert, 2011). The search direction can be significantly enhanced by multiplying the gradient with the inverse Hessian matrix, which serves as a deconvolution operator for compensating the geometrical spreading effects and de-blurring the gradient (Pratt et al., 1998). For multi-parameter FWI, the multi-parameter Hessian is also expected to suppress parameter crosstalk (or trade-off) (Operto et al., 2013; Innanen, 2014b; Pan et al., 2016b). Furthermore, the second-order term in the Hessian matrix which accounts for non-linear scattering effects, can help to remove the second-order scattering artifacts in the gradient (Pratt et al., 1998; Métivier et al., 2013; Pan et al., 2016b). However, explicit calculation, storage and inversion of the Hessian at each iteration is computationally impractical for large-scale inverse problems. Hence, various approaches have been proposed for approximating the Hessian (Shin et al., 2001; Plessix and Mulder, 2004; Tang, 2009; Jun et al., 2015) or inverse Hessian (Nocedal and Wright, 2006; Nammour and Symes, 2009; Demanet et al., 2012; Letourneau et al., 2012). In Gauss-Newton method, an approximate Hessian is introduced by involving only the first-order term and ignoring the second-order contributions (Pratt et al., 1998).

Instead of constructing the Hessian explicitly, quasi-Newton methods approximate the inverse Hessian iteratively by storing the model and gradient changes from previous iterations (Nocedal and Wright, 2006). One popular quasi-Newton method is the BFGS method (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970). However, the storage requirement of the inverse Hessian approximation and computation cost of preconditioning for large-scale inverse problems is still very high. To mitigate this difficulty, a limited-memory BFGS (l-BFGS) method is developed by storing information from a limited number l ($l < 10$) of previous iterations (Nocedal, 1980; Byrd et al., 1995; Nocedal and Wright, 2006). Compared to gradient-based methods, l-BFGS methods provide faster convergence rates for large-scale inverse problems (Brossier et al., 2010; Ma and Hale, 2012). The convergence performance of l-BFGS method is closely related to the initial guess of inverse Hessian approximation (Brossier et al., 2009; Guitton and Díaz, 2012).

Hessian-free optimization methods (truncated-Newton or inexact-Newton method) represent attractive alternatives to the above-described optimization methods (Nash, 1985; Santosa and Symes, 1988; Nash, 2000; Akcelik et al., 2002; Erlangga and Herrmann, 2009; Métivier et al., 2012; AlTheyab et al., 2013; Métivier et al., 2014). At each iteration,

the search direction is computed by approximately solving the Newton equations using a matrix-free fashion of the conjugate-gradient (CG) algorithm, which is an optimal method for solving a positive definite system (Nash, 1985; Hu et al., 2009). This linear iterative solver only requires the Hessian-vector products instead of forming the Hessian operator explicitly (Fichtner and Trampert, 2011; Métivier et al., 2014; Li and Demanet, 2016). In this research, the full Hessian is replaced with the Gauss-Newton Hessian, which is always symmetric and positive semi-definite. To accelerate the Hessian-free Gauss-Newton method, a l-BFGS inverse Hessian approximation is used to precondition the conjugate-gradient algorithm for improving the convergence rate and reduce the computational burden (Nash, 2000; Sainath et al., 2013; Pan et al., 2016a).

In multi-parameter FWI, the multi-parameter Hessian has a block structure. The off-diagonal blocks measure the parameter trade-offs between different physical parameters. Applying the inverse of the multi-parameter Hessian to precondition the gradient can reduce the parameter crosstalk artifacts. Hence, it is very necessary to apply Newton-based optimization methods for multi-parameter FWI.

In this paper, we first review the forward modelling problem in acoustic media with variable density and the basic principle of full-waveform inversion. Then, we give the sensitivity kernels for different parameterizations in multi-parameter acoustic FWI. The optimization methods including Newton-based methods, gradient-based methods, Quasi-Newton methods and Preconditioned Hessian-free optimization methods for FWI are explained. We also show how to construct the multi-parameter Hessian-vector products in Hessian-free Gauss-Newton methods for different parameterizations. In the numerical modelling section, we give numerical modelling results of the scattering patterns of different parameterizations in acoustic media. We also compare the inversion results of the different parameterizations with different optimization methods.

THEORY AND METHODS

Forward modelling in acoustic media with variable density

In the frequency domain, the forward modelling problem in acoustic medium with variable density is governed by the following equation (Marfurt, 1984):

$$\nabla \cdot \left(\frac{1}{\rho\left(\mathbf{x}\right)} \nabla u(\mathbf{x}, \mathbf{x}_s, \omega) \right) + \frac{\omega^2}{\kappa\left(\mathbf{x}\right)} u(\mathbf{x}, \mathbf{x}_s, \omega) = f_s(\omega) \delta\left(\mathbf{x} - \mathbf{x}_s\right), \tag{1}$$

where ω is the angular frequency, $\mathbf{x} = (x, y, z)$ denotes the subsurface location in Cartesian coordinates, ∇ is the first order spatial derivative operator, $\kappa = \rho v^2$ is the bulk modulus, ρ is the density, v is the velocity, $u(\mathbf{x}, \mathbf{x}_s, \omega)$ denotes the pressure wavefield at position $\mathbf{x}$, $\delta\left(\mathbf{x} - \mathbf{x}_s\right)$ is the Dirac delta function, and $f_s(\omega)$ means the source signature at position $\mathbf{x}_s$. In this software package, we develop 5-point and 9-point finite difference schemes to discretize the model (Jo et al., 1996) and a first-order Engquist-Majda boundary condition is implemented (Engquist and Majda, 1977). Equation (1) can be rewritten in a matrix form after discretization:

$$\mathbf{L}\left(\mathbf{m}, \omega\right) \mathbf{u}(\mathbf{x}, \mathbf{x}_s, \omega) = \mathbf{f}_s(\omega) \delta\left(\mathbf{x} - \mathbf{x}_s\right), \tag{2}$$

where $\mathbf{m}$ is the model parameter vector, $\mathbf{u}(\mathbf{x}, \mathbf{x}_s, \omega)$ and $\mathbf{f}_s(\omega)$ are the discrete pressure wavefield and source vectors. $\mathbf{L}(\mathbf{m}, \omega)$ is the discretized impedance matrix. The linear equation (equation (2)) is solved with a direct solver based on multi-frontal Lower Upper (LU) decomposition (Davis and Duff, 1997), which is efficient for a multi-source problem with forward and backward substitutions (Hu et al., 2011). The solution of equation (1) can be written as the convolution of source $f_s(\omega)$ with Green's function $G(\mathbf{x}, \mathbf{x}_s, \omega)$:

$$u(\mathbf{x}, \mathbf{x}_s, \omega) = f_s(\omega) G(\mathbf{x}, \mathbf{x}_s, \omega), \tag{3}$$

where the Green's function is defined as the solution of wave equation due to an impulse source.

Theory of full-waveform inversion: review

FWI allows to reconstruct high-resolution velocity models of the subsurface through the extraction of the full information content of the seismic data. The inversion process is implemented by minimizing a l-2 norm misfit function, which measures the difference between the modelled data and observed data:

$$\Phi(\mathbf{m}) = \frac{1}{2} \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_g} \sum_{\omega} \| \mathbf{d}_{\text{obs}}(\mathbf{x}_s, \mathbf{x}_g, \omega) - \mathbf{d}_{\text{syn}}(\mathbf{m}, \mathbf{x}_s, \mathbf{x}_g, \omega) \|^2, \tag{4}$$

where ω is the angular frequency, $\mathbf{m}$ indicates the model parameter vector, $\mathbf{x}_s$ and $\mathbf{x}_g$ are the source and receiver positions, $\mathbf{d}_{\text{obs}}$ and $\mathbf{d}_{\text{syn}} = \mathcal{P}\mathbf{u}$ indicate the observed data and synthetic data respectively, $\mathcal{P}$ is the sampling operator, and $\| \cdot \|$ means the l-2 norm. Here, we consider to minimize the objective function associated with the source weight:

$$\tilde{\Phi}(\mathbf{m}, \mathbf{s}) = \frac{1}{2} \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_g} \sum_{\omega} \| \mathbf{d}_{\text{obs}}(\mathbf{x}_s, \mathbf{x}_g, \omega) - \mathbf{s}(\mathbf{x}_s, \omega) \mathbf{d}_{\text{syn}}(\mathbf{m}, \mathbf{x}_s, \mathbf{x}_g, \omega) \|^2, \tag{5}$$

where $\mathbf{s}(\mathbf{x}_s, \omega)$ indicates the source weight vector.

The variable projection method is implemented by minimizing the objective function (equation (5)) with respect to both $\mathbf{m}$ and $\mathbf{s}$. We suppose that $\mathbf{d}_{\text{obs}}$ and $\mathbf{d}_{\text{syn}}$ are known, the source weight $\mathbf{s}$ can be obtained through a least-square approach:

$$\mathbf{s}(\mathbf{x}_s, \omega) = \frac{\sum_{\mathbf{x}_g} \mathbf{d}_{\text{obs}}(\mathbf{x}_s, \mathbf{x}_g, \omega) \mathbf{d}^*_{\text{syn}}(\mathbf{x}_s, \mathbf{x}_g, \omega)}{\sum_{\mathbf{x}_g} \mathbf{d}_{\text{syn}}(\mathbf{x}_s, \mathbf{x}_g, \omega) \mathbf{d}^*_{\text{syn}}(\mathbf{x}_s, \mathbf{x}_g, \omega)}, \tag{6}$$

where the symbol "*" means complex conjugate transpose. The Jacobian matrix is given as:

$$\mathbf{J} = \frac{\partial \mathbf{d}_{\text{syn}}}{\partial \mathbf{m}} = \mathcal{P}\mathbf{A}(\mathbf{m}, \omega)^{-1} \mathbf{A}_{\mathbf{m}}(\mathbf{m}, \omega) \mathbf{u} \tag{7}$$

where $\mathbf{A}_{\mathbf{m}}(\mathbf{m}, \omega) = \partial \mathbf{A}(\mathbf{m}, \omega) / \partial \mathbf{m}$. The gradient for minimizing objective function in equation (4) is obtained as:

$$\mathbf{g} = \frac{\partial \Phi(\mathbf{m})}{\partial \mathbf{m}} = \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_g} \sum_{\omega} \mathbf{u}^*(\mathbf{x}_s, \mathbf{x}_g, \omega) \mathbf{A}^*_{\mathbf{m}}(\mathbf{m}, \omega) \mathbf{A}^*(\mathbf{m}, \omega)^{-1} \mathcal{P}^* \Delta \mathbf{d}(\mathbf{x}_s, \mathbf{x}_g, \omega), \tag{8}$$

The approximate Hessian used in Gauss-Newton method is expressible as:

$$\mathbf{H}_a = \mathbf{J}^*\mathbf{J} = \mathbf{u}^*\mathbf{A}_\mathbf{m}^* \left(\mathbf{m},\omega\right)\mathbf{A}^* \left(\mathbf{m},\omega\right)^{-1}\mathcal{P}^*\mathcal{P}\mathbf{A} \left(\mathbf{m},\omega\right)^{-1}\mathbf{A}_\mathbf{m} \left(\mathbf{m},\omega\right)\mathbf{u}. \tag{9}$$

The gradient and Gauss-Newton Hessian can also be expressed using Green's functions. Considering acoustic media with constant density, the corresponding sensitivity kernel for the square of slowness can be expressed as:

$$K\left(\mathbf{x}\right) = -\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_r}\int_\omega\int_{\Omega(\mathbf{x})}\omega^2 u\left(\mathbf{x}_s,\mathbf{x},\omega\right)G^*\left(\mathbf{x},\mathbf{x}_r,\omega\right)d\mathbf{x}d\omega, \tag{10}$$

The Gauss-Newton Hessian can be expressed as:

$$\tilde{H}\left(\mathbf{x},\mathbf{x}'\right) = \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_r}\int_\omega\int_{\Omega(\mathbf{x}')}\int_{\Omega(\mathbf{x})} G\left(\mathbf{x}_s,\mathbf{x},\omega\right)G^*\left(\mathbf{x},\mathbf{x}_r,\omega\right)G\left(\mathbf{x}_s,\mathbf{x}',\omega\right)G^*\left(\mathbf{x}',\mathbf{x}_r,\omega\right)d\mathbf{x}d\mathbf{x}'d\omega, \tag{11}$$

In this section, we will show how to derive the sensitivity kernels for different parameterizations in acoustic media with variable density.

Sensitivity kernels of different parameterizations

We consider 4 parameterizations for acoustic media with variable density: (a) the bulk modulus-density parameterization; (b) the velocity-density parameterization; (c) the impedance-velocity parameterization; (d) the impedance-density parameterization. In this section, we give the expressions for the sensitivity kernels for different parameterizations in acoustic media. For bulk modulus-density parameterization, we define:

$$s_\kappa\left(\mathbf{x}\right) = \frac{1}{\kappa\left(\mathbf{x}\right)}, s_\rho\left(\mathbf{x}\right) = \frac{1}{\rho\left(\mathbf{x}\right)}. \tag{12}$$

Considering a model perturbation $\Delta s_\kappa\left(\mathbf{x}\right)$ at position $\mathbf{x}$, with Born approximation, the perturbed pressure wavefield can be obtained as:

$$\delta u\left(\mathbf{x}_r,\mathbf{x}_s,\omega\right) = -\omega^2\int_{\Omega(\mathbf{x})} u\left(\mathbf{x}_s,\mathbf{x},\omega\right)\Delta s_\kappa\left(\mathbf{x}\right)G\left(\mathbf{x},\mathbf{x}_r,\omega\right)d\mathbf{x}, \tag{13}$$

where $u\left(\mathbf{x}_s,\mathbf{x},\omega\right)$ and $G\left(\mathbf{x},\mathbf{x}_r,\omega\right)$ represent source-side wavefield and receiver-sides Green's function respectively. The corresponding sensitivity kernel can be written as:

$$K_\kappa\left(\mathbf{x}\right) = -\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_r}\int_\omega\int_{\Omega(\mathbf{x})}\omega^2 u\left(\mathbf{x}_s,\mathbf{x},\omega\right)\Delta s_\kappa\left(\mathbf{x}\right)u^*\left(\mathbf{x},\mathbf{x}_r,\omega\right)d\mathbf{x}d\omega, \tag{14}$$

where $u^*\left(\mathbf{x},\mathbf{x}_r,\omega\right)$ indicates the backpropagated residual wavefield:

$$u^*\left(\mathbf{x},\mathbf{x}_r,\omega\right) = \sum_{\mathbf{x}_s} G\left(\mathbf{x},\mathbf{x}_r,\omega\right)\Delta\mathbf{d}^*\left(\mathbf{x}_s,\mathbf{x}_r,\omega\right). \tag{15}$$

Similarly, when considering perturbation of density Δs, the corresponding sensitivity kernel can be obtained as:

$$K_\rho\left(\mathbf{x}\right) = \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_r}\int_\omega\int_{\Omega(\mathbf{x})}\nabla u\left(\mathbf{x}_s,\mathbf{x},\omega\right)\cdot\nabla u^*\left(\mathbf{x},\mathbf{x}_r,\omega\right)\Delta s_\rho\left(\mathbf{x}\right)d\mathbf{x}d\omega, \tag{16}$$

For velocity-density parameterization, we define:

$$s_\alpha(\mathbf{x}) = \frac{1}{\alpha(\mathbf{x})}, s'_\rho(\mathbf{x}) = \frac{1}{\rho(\mathbf{x})}. \tag{17}$$

To obtain the sensitivity kernels for velocity s_α and density s'_ρ in the velocity-density parameterization, we can employ the chain rule. First, we have:

$$s_\kappa(\mathbf{x}) = s'_\rho(\mathbf{x}) s^2_\alpha(\mathbf{x}), s'_\rho(\mathbf{x}) = s_\rho(\mathbf{x}). \tag{18}$$

Hence, we can obtain the sensitivity kernel for s_α by:

$$\begin{aligned} K_\alpha(\mathbf{x}) &= K_\rho(\mathbf{x}) \frac{\partial s_\kappa(\mathbf{x})}{\partial s_\alpha(\mathbf{x})} = 2 s'_\rho(\mathbf{x}) s_\alpha(\mathbf{x}) K_\rho(\mathbf{x}) = \\ &= 2 \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_r} \int_\omega \int_{\Omega(\mathbf{x})} s'_\rho(\mathbf{x}) s_\alpha(\mathbf{x}) \nabla u(\mathbf{x}_s, \mathbf{x}, \omega) \cdot \nabla u^*(\mathbf{x}, \mathbf{x}_r, \omega) \Delta s_\rho(\mathbf{x}) d\mathbf{x} d\omega. \end{aligned} \tag{19}$$

Similarly for the density sensitivity kernel, we have:

$$\begin{aligned} K'_\rho(\mathbf{x}) &= K_\rho(\mathbf{x}) + K_\kappa(\mathbf{x}) \frac{\partial s_\kappa(\mathbf{x})}{\partial s_\rho(\mathbf{x})} = K_\rho(\mathbf{x}) + s^2_\alpha(\mathbf{x}) K_\kappa(\mathbf{x}) \\ &= \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_r} \int_\omega \int_{\Omega(\mathbf{x})} \nabla u(\mathbf{x}_s, \mathbf{x}, \omega) \cdot \nabla u^*(\mathbf{x}, \mathbf{x}_r, \omega) \Delta s_\rho(\mathbf{x}) d\mathbf{x} d\omega \\ &- \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_r} \int_\omega \int_{\Omega(\mathbf{x})} \omega^2 s^2_\alpha(\mathbf{x}) u(\mathbf{x}_s, \mathbf{x}, \omega) \Delta s_\kappa(\mathbf{x}) u^*(\mathbf{x}, \mathbf{x}_r, \omega) d\mathbf{x} d\omega \end{aligned} \tag{20}$$

For impedance-density parameterization, we define:

$$s_{IP}(\mathbf{x}) = \frac{1}{IP(\mathbf{x})}, s''_\rho(\mathbf{x}) = \frac{1}{\rho(\mathbf{x})}, \tag{21}$$

where $IP = \alpha\rho$. We have:

$$s_\kappa(\mathbf{x}) = \frac{s^2_{IP}(\mathbf{x})}{s''_\rho(\mathbf{x})}, s''_\rho(\mathbf{x}) = s_\rho(\mathbf{x}). \tag{22}$$

According to the chain rule, we can derive the sensitivity kernels for impedance and density as:

$$\begin{aligned} K_{IP}(\mathbf{x}) &= \frac{\partial s_{\kappa(\mathbf{x})}}{\partial s_{IP}(\mathbf{x})} K_\kappa(\mathbf{x}) = \frac{2 s_{IP}(\mathbf{x})}{s''_\rho(\mathbf{x})} K_\kappa(\mathbf{x}) \\ &= -\sum_{\mathbf{x}_s} \sum_{\mathbf{x}_r} \int_\omega \int_{\Omega(\mathbf{x})} \frac{2 s_{IP}(\mathbf{x})}{s''_\rho(\mathbf{x})} \omega^2 u(\mathbf{x}_s, \mathbf{x}, \omega) \Delta s_\kappa(\mathbf{x}) u^*(\mathbf{x}, \mathbf{x}_r, \omega) d\mathbf{x} d\omega, \end{aligned} \tag{23}$$

$$\begin{aligned} K''_\rho(\mathbf{x}) &= K_\rho(\mathbf{x}) + \frac{\partial s_\kappa(\mathbf{x})}{\partial s''_\rho(\mathbf{x})} K_\kappa(\mathbf{x}) = K_\rho(\mathbf{x}) - \frac{s^2_{IP}(\mathbf{x})}{\left(s''_\rho\right)^2(\mathbf{x})} K_\kappa(\mathbf{x}) \\ &= -\sum_{\mathbf{x}_s} \sum_{\mathbf{x}_r} \int_\omega \int_{\Omega(\mathbf{x})} \frac{s^2_{IP}(\mathbf{x})}{\left(s''_\rho\right)^2(\mathbf{x})} \omega^2 u(\mathbf{x}_s, \mathbf{x}, \omega) \Delta s_\kappa(\mathbf{x}) u^*(\mathbf{x}, \mathbf{x}_r, \omega) d\mathbf{x} d\omega \\ &+ \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_r} \int_\omega \int_{\Omega(\mathbf{x})} \nabla u(\mathbf{x}_s, \mathbf{x}, \omega) \cdot \nabla u^*(\mathbf{x}, \mathbf{x}_r, \omega) \Delta s_\rho(\mathbf{x}) d\mathbf{x} d\omega. \end{aligned} \tag{24}$$

For impedance-velocity parameterization, we define:

$$s'_{IP}(\mathbf{x}) = \frac{1}{IP(\mathbf{x})}, s'_{\alpha}(\mathbf{x}) = \frac{1}{\alpha(\mathbf{x})}. \tag{25}$$

We have:

$$s_{\kappa}(\mathbf{x}) = s'_{IP}(\mathbf{x}) s'_{\alpha}(\mathbf{x}), s_{\rho}(\mathbf{x}) = \frac{s'_{IP}(\mathbf{x})}{s'_{\alpha}(\mathbf{x})}. \tag{26}$$

The sensitivity kernels for impedance and velocity are given by:

$$\begin{aligned} K'_{IP}(\mathbf{x}) &= \frac{\partial s_{\kappa}(\mathbf{x})}{\partial s'_{IP}(\mathbf{x})} K_{\kappa}(\mathbf{x}) + \frac{\partial s_{\rho}(\mathbf{x})}{\partial s_{IP}(\mathbf{x})} K_{\rho}(\mathbf{x}) = s'_{\alpha}(\mathbf{x}) K_{\kappa}(\mathbf{x}) + \frac{1}{s'_{\alpha}(\mathbf{x})} K_{\rho}(\mathbf{x}) \\ &= -\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_r}\int_{\omega}\int_{\Omega(\mathbf{x})} s'_{\alpha}(\mathbf{x})\omega^2 u(\mathbf{x}_s,\mathbf{x},\omega)\Delta s_{\kappa}(\mathbf{x}) u^*(\mathbf{x},\mathbf{x}_r,\omega)\, d\mathbf{x}d\omega \\ &+ \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_r}\int_{\omega}\int_{\Omega(\mathbf{x})} \frac{1}{s'_{\alpha}(\mathbf{x})}\nabla u(\mathbf{x}_s,\mathbf{x},\omega)\cdot\nabla u^*(\mathbf{x},\mathbf{x}_r,\omega)\Delta s_{\rho}(\mathbf{x})\, d\mathbf{x}d\omega, \end{aligned} \tag{27}$$

$$\begin{aligned} K'_{\alpha}(\mathbf{x}) &= \frac{\partial s_{\kappa}(\mathbf{x})}{\partial s'_{\alpha}(\mathbf{x})} K_{\kappa}(\mathbf{x}) + \frac{\partial s_{\rho}(\mathbf{x})}{\partial s_{\alpha}(\mathbf{x})} K_{\rho}(\mathbf{x}) = s'_{IP}(\mathbf{x}) K_{\kappa}(\mathbf{x}) - \frac{s'_{IP}(\mathbf{x})}{(s'_{\alpha})^2(\mathbf{x})} K_{\rho}(\mathbf{x}) \\ &= -\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_r}\int_{\omega}\int_{\Omega(\mathbf{x})} s'_{IP}(\mathbf{x})\omega^2 u(\mathbf{x}_s,\mathbf{x},\omega)\Delta s_{\kappa}(\mathbf{x}) u^*(\mathbf{x},\mathbf{x}_r,\omega)\, d\mathbf{x}d\omega \\ &- \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_r}\int_{\omega}\int_{\Omega(\mathbf{x})} \frac{s'_{IP}(\mathbf{x})}{(s'_{\alpha})^2(\mathbf{x})}\nabla u(\mathbf{x}_s,\mathbf{x},\omega)\cdot\nabla u^*(\mathbf{x},\mathbf{x}_r,\omega)\Delta s_{\rho}(\mathbf{x})\, d\mathbf{x}d\omega. \end{aligned} \tag{28}$$

Optimization methods

Full-Newton and Gauss-Newton methods

Newton-type optimization methods (e.g., full Newton (FN) and Gauss-Newton (GN) methods) use the quadratic search direction and exhibit fast convergence given a limited number of unknown parameters. The FN search direction is formed by preconditioning the gradient with the full Hessian $\mathbf{H}$: $\Delta\mathbf{m}_k = -\mathbf{H}_k^{-1}\mathbf{g}_k$. The Gauss-Newton approximate Hessian $\tilde{\mathbf{H}}$ only accounts for the first-order scattering effects. The element $\tilde{\mathbf{H}}(\mathbf{x},\mathbf{x}')$ is formed by correlating two Fréchet derivative wavefields at the receivers' locations due to model perturbations at positions $\mathbf{x}$ and $\mathbf{x}'$. Because of the band-limited signature of the seismic data, the Gauss-Newton Hessian $\tilde{\mathbf{H}}$ is diagonally dominant and banded (Pratt et al., 1998; Valenciano, 2008; Tang, 2009; Pan et al., 2014). The diagonal elements compensate for the geometrical spreading and the off-diagonal elements are responsible for de-blurring the gradient (Pratt et al., 1998). In multi-parameter FWI, the Gauss-Newton Hessian can mitigate the parameter crosstalk problem (Operto et al., 2013; Innanen, 2014a; Pan et al., 2016b).

For these Newton-type methods, explicit evaluation and inversion the Hessian matrix $\mathbf{H}$ and Gauss-Newton Hessian $\tilde{\mathbf{H}}$ at each iteration are required. Though Newton-type methods benefit from fast convergence rate, the computation, storage and inversion of Hessian at each iteration are prohibitively expensive, which limits their applications for large-scale inverse problems in exploration geophysics.

Gradient-based methods

Gradient-based methods (e.g., steepest-descent (SD) and non-linear conjugate-gradient (NCG) methods) approximate the Hessian matrix $\mathbf{H}$ as an identity matrix $\mathbf{I}$ and they are computationally more attractive than the Newton-type ones when inverting a large number of unknown model parameters. The SD method simply determines the search direction to be the negative of the gradient. In NCG method, the search direction is just a linear combination of current gradient and previous search direction. The gradient-based methods are known to converge globally, but possibly very slowly. In most cases, preconditioning is necessary to ensure the fast convergence of the CG method (Hu et al., 2011).

Quasi-Newton methods

Quasi-Newton methods provide an attractive alternative to Newton-type and gradient-based methods by approximating the inversion Hessian iteratively instead of constructing the Hessian matrix explicitly (Brossier et al., 2009; Ma and Hale, 2012). BFGS method, named after Broyden (1970), Fletcher (1970), Goldfarb (1970) and Shanno (1970), is one popular quasi-Newton strategy to approximate the inverse Hessian iteratively using the changes of the model and gradient (Nocedal and Wright, 2006).

In the BFGS updating formula, we are given a symmetric and positive definite matrix $\mathcal{H}_k$ that approximates the inverse of the Hessian, and a pair of vectors $\mathbf{s}_k = \mathbf{m}_{k+1} - \mathbf{m}_k$, and $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ that indicates the model and gradient changes and satisfies the condition $\mathbf{s}_k^\dagger \mathbf{y}_k > 0$. Using these vectors, we compute the inverse Hessian approximation $\mathcal{H}_{k+1}$ by the following formula:

$$\mathcal{H}_{k+1} = \mathbf{v}_k^\dagger \mathcal{H}_k \mathbf{v}_k + \mathbf{w}_k \mathbf{s}_k \mathbf{s}_k^\dagger, \tag{29}$$

where $\mathbf{w}_k = 1/\mathbf{y}_k^\dagger \mathbf{s}_k$, $\mathbf{v}_k = \mathbf{I} - \mathbf{w}_k \mathbf{y}_k \mathbf{s}_k^\dagger$ and $\mathbf{I}$ is the identity matrix. The initial inverse Hessian approximation $\mathcal{H}_0$ is important to BFGS method and it is usually set as an identity matrix to make sure that the updated matrix maintains positive definiteness (Wu et al., 2015). A limited-memory BFGS (l-BFGS) method was developed by storing the model and gradient changes from a limited number l of previous iterations (typically $l < 10$) (Nocedal, 1980). The stored information is then used to construct an approximated inverse Hessian. A "two-loop recursion" scheme is implemented in this research to obtain the search direction using the information of previous updates (Nocedal and Wright, 2006; Métivier and Brossier, 2016).

l-BFGS preconditioned Hessian-free optimization methods

Instead of constructing Hessian or inverse Hessian approximations, the Hessian-free (HF) optimization method, also known as truncated-Newton or inexact-Newton method,

obtains the search direction by solving the Newton linear system (equation (**??**)) approximately using a conjugate-gradient (CG) method with matrix-free scheme (Saad, 2003; Anagaw and Sacchi, 2012; Métivier et al., 2014). The CG method is an optimal algorithm for solving a symmetric positive definite system **Wx=b** and it only requires computing the Hessian-vector products $\mathbf{H}\upsilon$ instead of forming the Hessian matrix explicitly, where υ is an arbitrary vector in model space. The Hessian-vector products can be calculated via finite-difference method (Nocedal and Wright, 2006) or the second-order adjoint-state method (Fichtner and Trampert, 2011; Métivier et al., 2014; Métivier and Brossier, 2016). In this paper, the second method is employed for the calculation of the Hessian-vector products.

The full Hessian $\mathbf{H}$ arising from the second-order partial derivative is not guaranteed to be positive definite (Nash, 2000). Thus, CG method is no longer appropriate for solving an indefinite linear system. In this paper, instead of using the Hessian $\mathbf{H}$, we choose Gauss-Newton Hessian $\tilde{\mathbf{H}}$, which is always symmetric and positive semi-definite:

$$\left(\tilde{\mathbf{H}}_k + \epsilon\hat{\mathbf{A}}_k\right)\Delta\mathbf{m}_k = -\mathbf{g}_k, \tag{30}$$

where $\epsilon\hat{\mathbf{A}}_k$ is the damping term ensuring that $\tilde{\mathbf{H}}_k + \epsilon\hat{\mathbf{A}}$ is positive definite, ϵ is a small constant value and $\hat{\mathbf{A}}_k$ indicates a diagonal matrix consisting of the diagonal elements of the Gauss-Newton Hessian. The resulting algorithm becomes a Levenberg-Marquardt method (Levenberg, 1944; Marquardt, 1963). The HF Gauss-Newton FWI is implemented in a double-iterative scheme: the outer loop is to iteratively update the model parameters for the non-linear optimization problem, and the inner loop is to solve the linear system iteratively with the CG algorithm. The inner iteration is typically stopped or “truncated" before the solution of the Newton equation is obtained. A Hessian-free optimization method can be made more competitive with further enhancements, such as, an effective preconditioner for the linear system and appropriate stopping criteria for the inner iterative algorithm. With these enhancements, Hessian-free optimization method is a powerful tool for large-scale inverse problems.

The CG iterative algorithm requires many iterations to obtain the approximate solution of a linear system $\mathbf{Wx} = \mathbf{b}$. The convergence rate of the CG method depends on the spectral properties (e.g., its eigenvalues) of the coefficient matrix $\mathbf{W}$ (Nash, 2000). It is often convenient to transform the equation system into one which has the same solution but more favorable spectral properties. This can be achieved by applying a suitable preconditioner $\mathcal{M}$ on the linear system: $\mathcal{M}^{-1}\mathbf{Wx} = \mathcal{M}^{-1}\mathbf{b}$. Thus, the preconditioned Newton system for the HF Gauss-Newton FWI is given by:

$$\mathcal{M}_k^{-1}\left(\tilde{\mathbf{H}}_k + \epsilon\hat{\mathbf{A}}_k\right)\Delta\mathbf{m}_k = -\mathcal{M}_k^{-1}\mathbf{g}_k. \tag{31}$$

The solution of equation (31) can be obtained by the preconditioned conjugate-gradient (PCG) method. The preconditioner for the CG method is always devised to approximate the Hessian or the inverse Hessian. In this research, we develop an l-BFGS preconditioning scheme for the HF optimization method (Nash, 1985, 2000; Métivier et al., 2012; Sainath et al., 2013), namely l-BFGS-GN method for $\mathcal{H}_0 = \mathbf{I}$. The l-BFGS approximated inverse Hessian $\mathcal{H}$ can also be used as a preconditioner for the CG iterative method:

$$\mathcal{H}_k\left(\tilde{\mathbf{H}}_k + \epsilon\hat{\mathbf{A}}_k\right)\Delta\mathbf{m}_k = -\mathcal{H}_k\mathbf{g}_k. \tag{32}$$

Traditionally, an identity matrix $\mathbf{I}$ is usually set as the initial guess $\mathcal{H}_0$.

Multi-parameter Gauss-Newton Hessian for different parameterizations

The role of the single-parameter Hessian has been discussed and analyzed in the previous section. Multi-parameter Hessian in multi-parameter FWI has a block structure and it carries more information than the single-parameter Hessian. Considering a 2D subsurface model with $N_x N_z$ nodes and N_p physical parameters are assigned to describe the properties of each node. The multi-parameter Hessian is a $N_x N_z N_p \times N_x N_z N_p$ square and symmetric matrix with N_p diagonal blocks and $N_p\left(N_p-1\right)$ off-diagonal blocks. Each block is a $N_x N_z \times N_x N_z$ square matrix. The multi-parameter Hessian $\mathbf{H}$ can be written as the summation of the first-order term $\tilde{\mathbf{H}}$ and second-order term $\bar{\mathbf{H}}$. The first-order term $\tilde{\mathbf{H}}$ is also known as multi-parameter approximate Hessian used in Gauss-Newton method. The elements in $\tilde{\mathbf{H}}$ measure the correlations of two Fréchet derivative wavefields. For example, the element $\tilde{\mathbf{H}}_{\mathbf{m}_1\mathbf{m}_2}\left(\mathbf{x}, \mathbf{x}'\right)$ can be expressed as:

$$\tilde{\mathbf{H}}_{\mathbf{m}_1\mathbf{m}_2}\left(\mathbf{x}, \mathbf{x}'\right)=\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\sum_{\omega}\Re\left(\frac{\partial \mathbf{u}^{\dagger}\left(\mathbf{x}_g, \mathbf{x}_s, \omega\right)}{\partial \mathbf{m}_1\left(\mathbf{x}\right)}\frac{\partial \mathbf{u}^{*}\left(\mathbf{x}_g, \mathbf{x}_s, \omega\right)}{\partial \mathbf{m}_2\left(\mathbf{x}'\right)}\right), \tag{33}$$

where when $\mathbf{m}_1=\mathbf{m}_2$, it indicates the element in diagonal block, and when $\mathbf{m}_1 \neq \mathbf{m}_2$, it indicates the element in off-diagonal block. The multi-parameter approximate Hessian $\tilde{\mathbf{H}}$ is essential in overcoming the crosstalk difficulty in multi-parameter FWI (Operto et al., 2013; Pan et al., 2015). As we discussed in the previous section, the similarity of the Fréchet derivative wavefields with respect to different physical parameters gives rise to the crosstalk problem. The off-diagonal blocks in multi-parameter approximate Hessian predict the coupling effects and applying its inverse to the gradient can remove or mitigate the parameter crosstalk. The space-type multi-parameter Hessian approximation $\tilde{\mathbf{H}}_{\mathbf{s}}$ given by Innanen (2014a) neglects the contributions of the off-diagonal blocks and stresses the correlation of Fréchet derivative wavefields with respect to the same physical parameter. This approximation can scale the amplitudes of the gradient and de-blur the gradient, but can not suppress parameter crosstalk. The parameter-type multi-parameter Hessian approximation $\tilde{\mathbf{H}}_{\mathbf{p}}$ only keeps the diagonal elements of the blocks, which is also capable of mitigating parameter crosstalk but limited in resolving the gradient.

Here, we give the multi-parameter Gauss-Newton Hessian expressions for different parameterizations in acoustic media. First, considering the velocity-density parameterization, the multi-parameter Gauss-Newton Hessian are correlation of two Fréchet derivative wave-

fields:

$$\begin{aligned}\tilde{H}_{\text{vel}}(\mathbf{x},\mathbf{x}') &= \begin{bmatrix} K_\alpha(\mathbf{x}) \\ K_\rho(\mathbf{x}) \end{bmatrix} \begin{bmatrix} K_\alpha^\dagger(\mathbf{x}') & K_\rho^\dagger(\mathbf{x}') \end{bmatrix} \\ &= \begin{bmatrix} K_\alpha(\mathbf{x}) K_\alpha^\dagger(\mathbf{x}') & K_\alpha(\mathbf{x}) K_\rho^\dagger(\mathbf{x}') \\ K_\rho(\mathbf{x}) K_\alpha^\dagger(\mathbf{x}') & K_\rho(\mathbf{x}) K_\rho^\dagger(\mathbf{x}') \end{bmatrix} \\ &= \begin{bmatrix} \tilde{H}_{\alpha\alpha}(\mathbf{x},\mathbf{x}') & \tilde{H}_{\alpha\rho}(\mathbf{x},\mathbf{x}') \\ \tilde{H}_{\rho\alpha}(\mathbf{x},\mathbf{x}') & \tilde{H}_{\rho\rho}(\mathbf{x},\mathbf{x}') \end{bmatrix},\end{aligned} \tag{34}$$

Similarly, for impedance-velocity and impedance-density parameterizations, the multi-parameter Gauss-Newton Hessian can be written as:

$$\begin{aligned}\tilde{H}_{\text{imp}}(\mathbf{x},\mathbf{x}') &= \begin{bmatrix} K_{IP}(\mathbf{x}) \\ K'_\rho(\mathbf{x}) \end{bmatrix} \begin{bmatrix} K_{IP}^\dagger(\mathbf{x}') & (K')_\rho^\dagger(\mathbf{x}') \end{bmatrix} \\ &= \begin{bmatrix} K_{IP}(\mathbf{x}) K_{IP}^\dagger(\mathbf{x}') & K_{IP}(\mathbf{x}) (K')_\rho^\dagger(\mathbf{x}') \\ (K')_\rho(\mathbf{x}) K_{IP}^\dagger(\mathbf{x}') & K'_\rho(\mathbf{x}) (K')_\rho^\dagger(\mathbf{x}') \end{bmatrix} \\ &= \begin{bmatrix} \tilde{H}_{IPIP}(\mathbf{x},\mathbf{x}') & \tilde{H}_{IP\rho}(\mathbf{x},\mathbf{x}') \\ \tilde{H}_{\rho IP}(\mathbf{x},\mathbf{x}') & \tilde{H}_{\rho\rho}(\mathbf{x},\mathbf{x}') \end{bmatrix},\end{aligned} \tag{35}$$

$$\begin{aligned}\tilde{H}_{\text{imp}}(\mathbf{x},\mathbf{x}') &= \begin{bmatrix} K'_{IP}(\mathbf{x}) \\ K'_\alpha(\mathbf{x}) \end{bmatrix} \begin{bmatrix} (K')_{IP}^\dagger(\mathbf{x}') & (K')_\alpha^\dagger(\mathbf{x}') \end{bmatrix} \\ &= \begin{bmatrix} (K')_{IP}(\mathbf{x}) (K')_{IP}^\dagger(\mathbf{x}') & (K')_{IP}(\mathbf{x}) (K')_\rho^\dagger(\mathbf{x}') \\ (K')_\rho(\mathbf{x}) (K')_{IP}^\dagger(\mathbf{x}') & K'_\rho(\mathbf{x}) (K')_{alpha}^\dagger(\mathbf{x}') \end{bmatrix} \\ &= \begin{bmatrix} \tilde{H}_{IPIP}(\mathbf{x},\mathbf{x}') & \tilde{H}_{IP\alpha}(\mathbf{x},\mathbf{x}') \\ \tilde{H}_{\alpha IP}(\mathbf{x},\mathbf{x}') & \tilde{H}_{\alpha\alpha}(\mathbf{x},\mathbf{x}') \end{bmatrix}.\end{aligned} \tag{36}$$

NUMERICAL EXPERIMENTS

In this section, we first show the numerical modelling results of the scattering patterns of different parameterizations are examined. Figures 1a and 1b show the scattering patterns of velocity and density for velocity-density parameterization. Figures 1c and 1d show the scattering patterns of impedance and density for impedance-density parameterization. Figures 1e and 1f show the scattering patterns of impedance and density for impedance-velocity parameterization. As we can see, different parameterizations give different scattering patterns indicating different resolving abilities of these parameterizations.

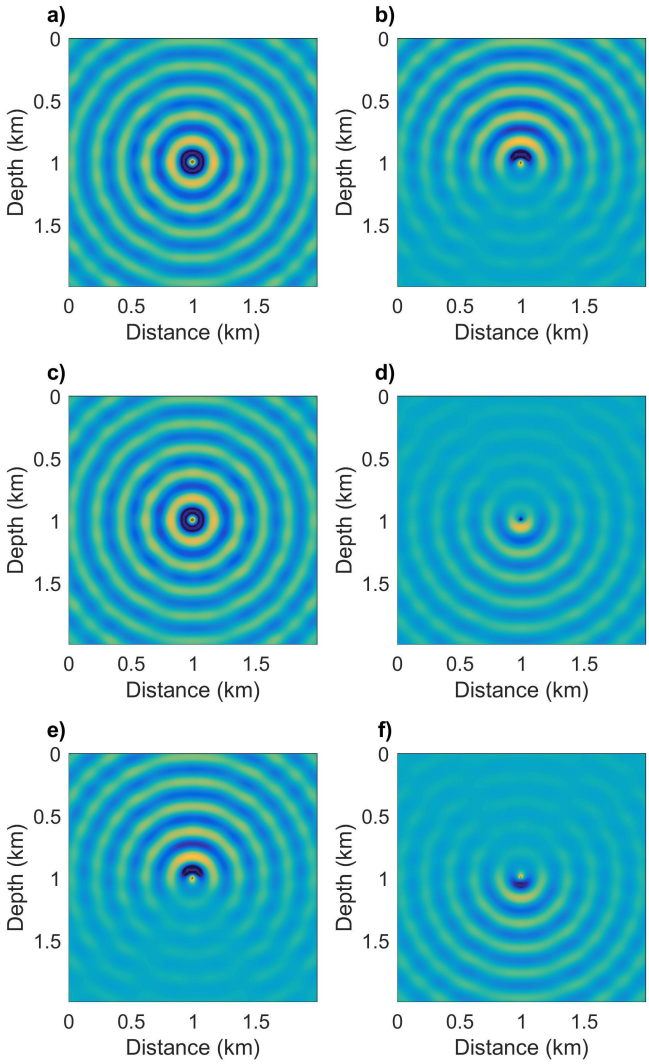

FIG. 1. (a) and (b) show the scattering patterns of velocity and density for the velocity-density parameterization; (c) and (d) show the scattering patterns of impedance and density for the impedance-density parameterization; (e) and (f) show the scattering patterns of impedance and velocity for the impedance-velocity parameterization.

In theory, we find that for the velocity-density parameterization with reflection survey, these two parameters will be difficult to reconstruct at short apertures. However, with transmission survey, the mapping from P-wave velocity to density will be very strong and

the mapping form density to velocity will be very weak. This means that for transmission survey, it will be easier to recover velocity but more difficult to recover density due to the strong crosstalk from velocity. Similarly, for impedance-density and impedance-velocity parameterizations, we predict that it will

There are other problems when employing impedance-velocity or impedance-density parameterization. Because the magnitudes velocity and density are much smaller than the magnitude of impedance. It will not be appropriate to reconstruct impedance and density or impedance and velocity simultaneously when using the same step length to scale the updates. A hierarchical strategy is better to be used

We next give an example to show that the multi-parameter Hessian can help to reduce the parameter crosstalk difficulty in multi-parameter FWI. Figures 2a and 2b show the true P-wave velocity and density model parameters. The true model consists of two uncorrelated Gaussian anomalies. Figures 3a and 3b show the gradient updates for P-wave velocity and density respectively. As we can see, the mapping from density to P-wave velocity is very weak. While the mapping from P-wave velocity to density is very strong, which make density difficult to reconstruct for transmission survey. Figure Figures 4a and 4b show the Gauss-Newton updates for P-wave velocity and density with 5 inner conjugate-gradient iterations. We observe that the spatial resolution of the model updates have been improved. Furthermore, in the density update, the crosstalk artifacts are suppressed and the true update for density has been enhanced, which indicates that the Hessian can help to reduce the parameter crosstalk for multi-parameter FWI.

Figures 5a and 5b show the inverted P-wave velocity and density models using steepest descent (SD) method after 10 iterations. It can be seen that inverted models are far from the true models because of the slow convergency rate of SD method. Figures 6a and 6b show the inverted P-wave velocity and density models using l-BFGS method after 10 iterations. As we can see, the inverted models have been enhanced obviously. However, the parameter crosstalk artifacts in the inverted density model is still very strong. Figures 7a and 7b show the inverted P-wave velocity and density models using non-preconditioned Hessian-free Gauss-Newton method after 10 iterations. We see that the parameter crosstalk artifacts in the inverted density model has been obviously suppressed. Figures 8a and 8b show the inverted P-wave velocity and density models using l-BFGS preconditioned Hessian-free Gauss-Newton method after 10 iterations. With preconditioning, we see that the parameter crosstalk artifacts are suppressed further. Figure 9 shows the convergence history, P-wave velocity errors, and density errors. The l-BFGS preconditioned Hessian-free Gauss-Newton method can reconstruct the models better than gradient-based and l-BFGS methods.

In the next example, we illustrate the inverted models using different optimization methods with a more complex model. Figures 11a and 11c show the true P-wave velocity and density models. Figure 11c and 11d show the initial P-wave velocity and density models. A transmission survey is employed for inversion. Figures 11a and 11b show the inverted velocity and density models using SD method. Figures 11c and 11d show the inverted velocity and density models using NCG method. Figures 11e and 11f show the inverted velocity and density models using l-BFGS method. Figures 11g and 11h show the inverted

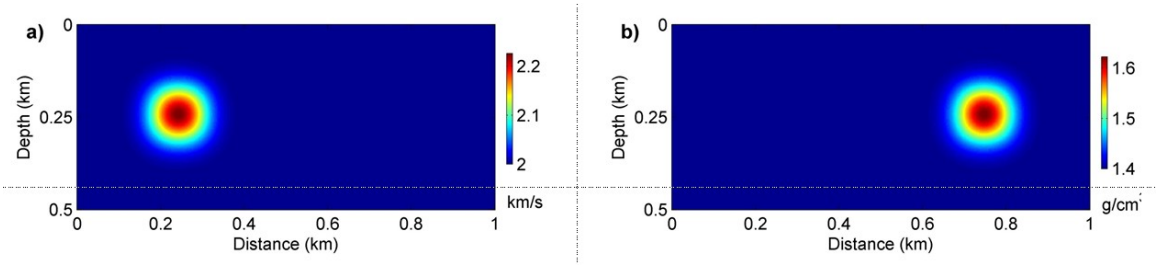

FIG. 2. (a) and (b) show the true velocity and density models.

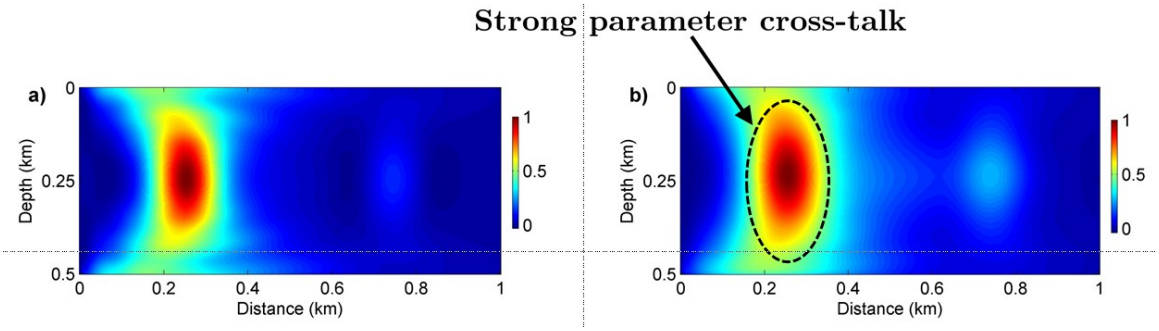

FIG. 3. (a) and (b) show the gradient updates for P-wave velocity and density.

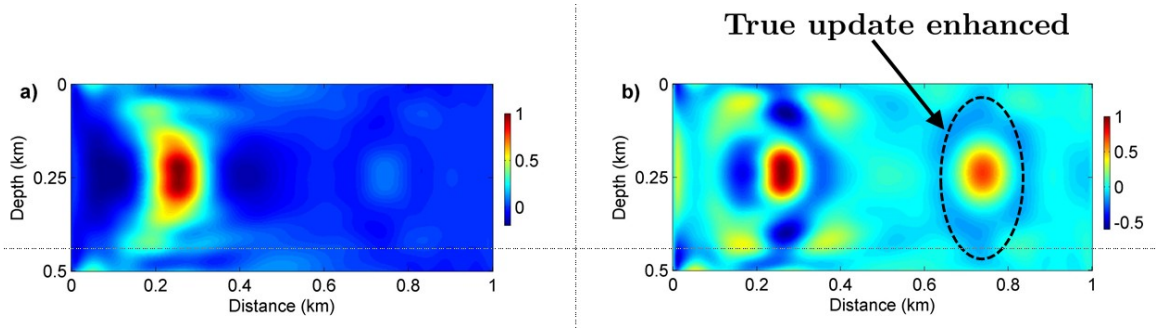

FIG. 4. (a) and (b) show the Gauss-Newton updates for P-wave velocity and density.

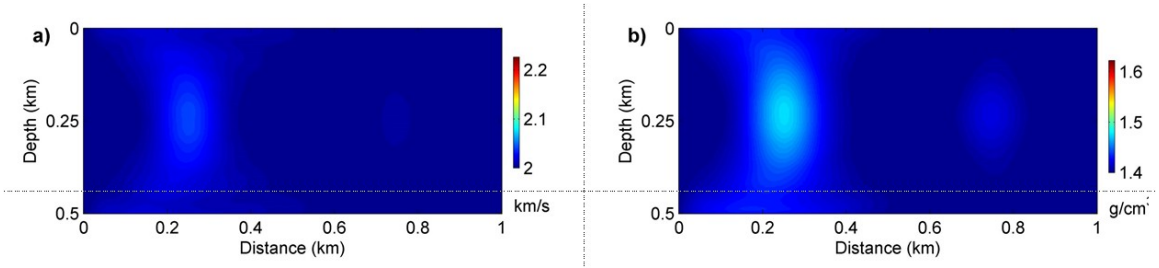

FIG. 5. (a) and (b) show the inverted P-wave velocity and density models with SD method.

velocity and density models using non-preconditioned HFGN method. Figures 11i and 11j show the inverted velocity and density models using l-BFGS preconditioned HFGN method. We see that the preconditioned HFGN method gives best inverted models.

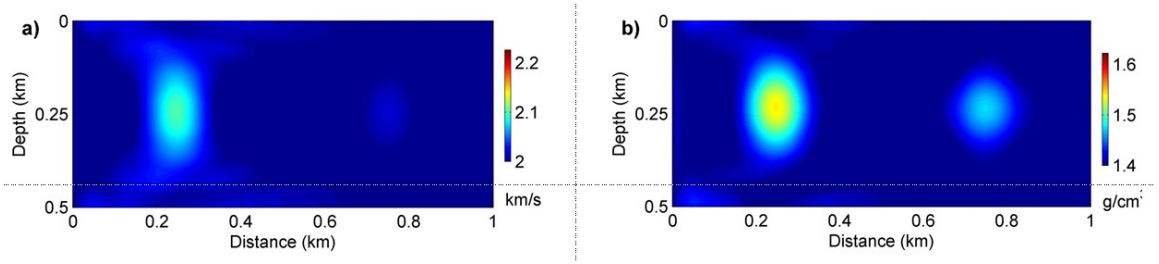

FIG. 6. (a) and (b) show the inverted P-wave velocity and density models with LBFGS method.

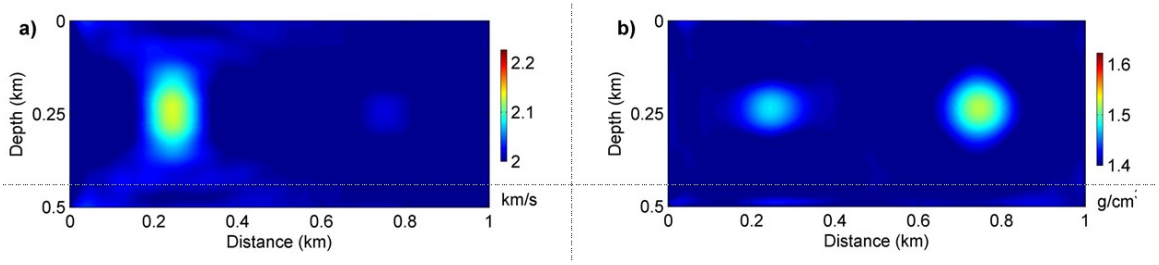

FIG. 7. (a) and (b) show the inverted P-wave velocity and density models with non-preconditioned Hessian-free Gauss-Newton method.

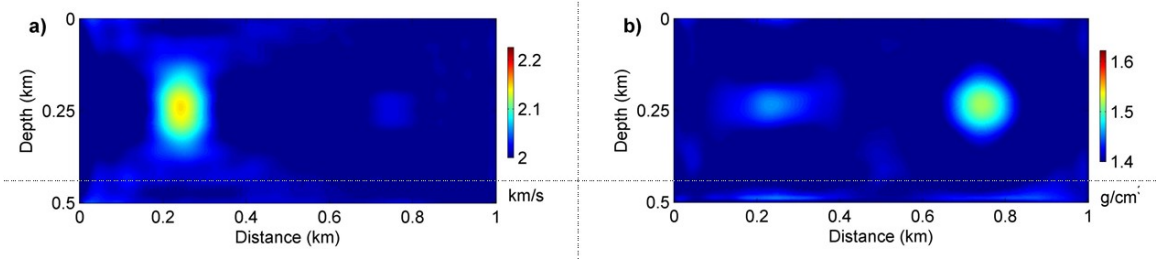

FIG. 8. (a) and (b) show the inverted P-wave velocity and density models with l-BFGS preconditioned Hessian-free Gauss-Newton method.

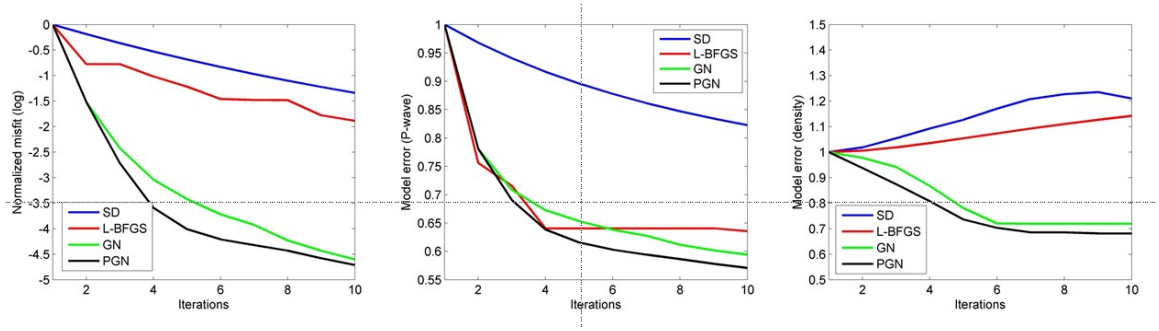

FIG. 9. The left, middle and right figures show the convergency history, P-wave velocity errors, and density errors as iteration proceeds.

CONCLUSION

In this research, we give the expressions of the sensitivity kernels for different parameterizations in multi-parameter acoustic FWI. We also examine the scattering patterns for different parameterizations. Numerical example is given to show the effectiveness of the multi-parameter Hessian in suppressing the parameter crosstalk artifacts. Different optimizations are also compared for reconstructing velocity and density simultaneously.

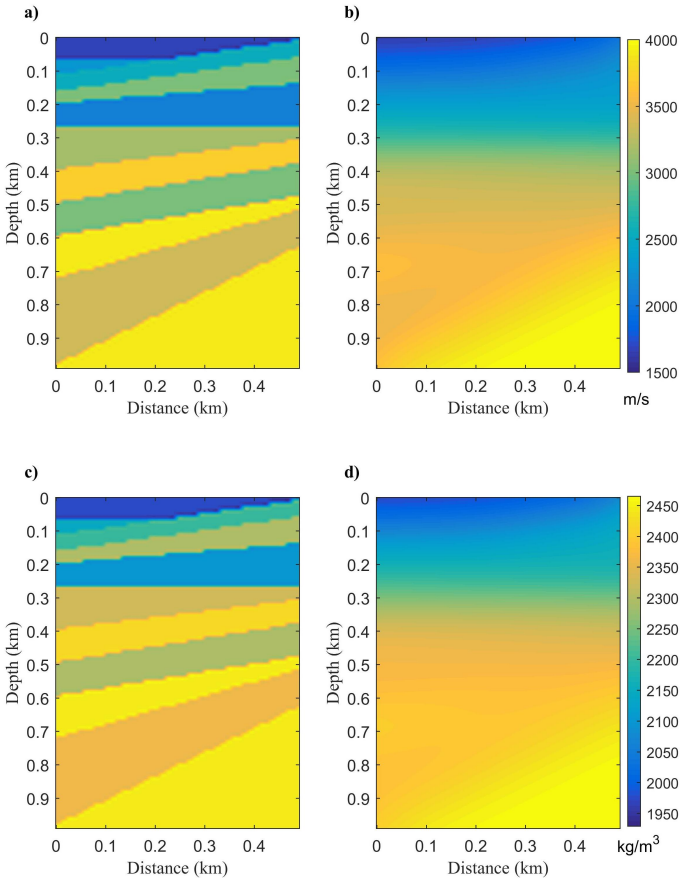

FIG. 10. (a) and (b) show the true velocity and density models. (c) and (d) show the initial velocity and density models.

ACKNOWLEDGEMENTS

The authors thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13. Author 1 was also supported by a SEG/Chevron scholarship and eye high international doctoral scholarship.

REFERENCES

Akcelik, V., Biros, G., and Ghattas, O., 2002, Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation: Proceedings of the ACM/IEEE Conference on Supercomputing, 2357–2361.

AlTheyab, A., Wang, X., and Schuster, G. T., 2013, Time-domain incomplete Gauss-Newton full-waveform inversion of Gulf of Mexico data: SEG Technical Program Expanded Abstracts, 5175–5179.

Anagaw, A. Y., and Sacchi, M. D., 2012, Full waveform inversion with simultaneous sources using the full Newton method: SEG Expanded Abstracts, 971–975.

Brossier, R., Operto, S., and Virieux, J., 2009, Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion: Geophysics, **74**, WCC105–WCC118.

Brossier, R., Operto, S., and Virieux, J., 2010, Which data residual norm for robust elastic frequency-domain full-waveform inversion?: Geophysics, **75**, R37–R46.

Broyden, C. G., 1970, The convergence of a class of double-rank minimization algorithms: IMA Journal of Applied Mathematics, **6**, 222–231.

Byrd, R. H., Lu, P., and Nocedal, J., 1995, A limited memory algorithm for bound constrained optimization: SIAM Journal on Scientific and Statistical Computing, **16**, 1190–1208.

Davis, T. A., and Duff, I. S., 1997, An unsymmetric pattern multifrontal method for sparse lu factorization: SIAM Journal on Matrix Analysis and Applications, **18**, 140–158.

Demanet, L., Letourneau, P., Boumal, N., Calandra, H., and Snelson, S., 2012, Matrix probing: a randomized preconditioner for the wave equation Hessian: Applied and Computational Harmonic Analysis, **32**, R25–R36.

Engquist, B., and Majda, A., 1977, Absorbing boundary conditions for the numerical simulations of waves: Mathematical Computation, **31**, 629–651.

Erlangga, Y. A., and Herrmann, F. J., 2009, Seismic waveform inversion with Gauss-Newton-Krylov method: SEG Technical Program Expanded Abstracts, 2357–2361.

Fichtner, A., and Trampert, J., 2011, Hessian kernels of seismic data functionals based upon adjoint techniques: Geophysical Journal International, **185**, 775–798.

Fletcher, R., 1970, A new approach to variable metric algorithms: Computer Journal, **13**, 317–322.

Goldfarb, D., 1970, A family of variable-metric methods derived by variational means: Mathematics of Computation, **24**, 23–26.

Guitton, A., and Díaz, E., 2012, Attenuating crosstalk noise with simultaneous source full waveform inversion: Geophysical Prospecting, **60**, 759–768.

Hu, W., Abubakar, A., and Habashy, T. M., 2009, Simultaneous multifrequency inversion of full-waveform seismic data: Geophysics, **74**, R1–R14.

Hu, W., Abubakar, A., Habashy, T. M., and Liu, J., 2011, Preconditioned non-linear conjugate gradient method for frequency domain full-waveform seismic inversion: Geophysical Prospecting, **59**, 477–491.

Innanen, K. A., 2014a, Reconciling seismic AVO and precritical reflection FWI-analysis of the inverse Hessian: SEG Technical Program Expanded Abstracts, 1022–1027.

Innanen, K. A., 2014b, Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion: Geophysical Journal International, **199**, 717–734.

Jo, C. H., Shin, C., and Suh, J. H., 1996, An optimal 9-point, finite-difference, frequency-space, 2-D scalar wave extrapolator: Geophysics, **61**, 529–537.

Jun, H., Park, E., and Shin, C., 2015, Weighted pseudo-Hessian for frequency-domain elastic full waveform inversion: Journal of Applied Geophysics, **123**, 1–17.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migration: Conference on Inverse Scattering, Theory and Applications, SIAM, Expanded Abstracts, 206–220.

Letourneau, P., Demanet, L., and Calandra, H., 2012, Approximate inversion of the wave-equation Hessian via randomized matrix probing: SEG Technical Program Expanded Abstracts, 1–6.

Levenberg, K., 1944, A method for the solution of certain non-linear problems in least squares: Quarterly of Applied Mathematics, **2**, 164–168.

Li, Y. E., and Demanet, L., 2016, Full-waveform inversion with extrapolated low-frequency data: Geophysics, **81**, R339–R348.

Ma, Y., and Hale, D., 2012, Quasi-Newton full-waveform inversion with a projected Hessian matrix: Geophysics, **77**, R207–R216.

Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: Geophysics, **49**, 533–549.

Marquardt, D., 1963, An algorithm for least-squares estimation of nonlinear parameters: SIAM Journal, **11**, 431–441.

Métivier, L., Bretaudeau, F., Brossier, R., Virieux, J., and Operto, S., 2014, Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures: Geophysical Prospecting, **62**, 1–23.

Métivier, L., and Brossier, R., 2016, The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication: Geophysics, **81**, F1–F15.

Métivier, L., Brossier, R., Virieux, J., and Operto, S., 2012, The truncated Newton method for full waveform inversion: SEG Technical Program Expanded Abstracts, 1–5.

Métivier, L., Brossier, R., Virieux, J., and Operto, S., 2013, Full waveform inversion and the truncated Newton method: SIAM Journal On Scientific Computing, **35**, B401–B437.

Nammour, R., and Symes, W., 2009, Approximate constant density acoustic inverse scattering using dip-dependent scaling: SEG Technical Program Expanded Abstracts, 2347–2351.

Nash, S. G., 1985, Preconditioning of truncated-Newton methods: SIAM Journal on Scientific and Statistical Computing, **6**, 599–616.

Nash, S. G., 2000, A survey of truncated-Newton methods: Journal of Computational and Applied Mathematics, **124**, 45–59.

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–782.

Nocedal, J., and Wright, S. J., 2006, Numerical Optimization: Springer.

Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L., and Virieux, J., 2013, A guided tour of multiparameter full waveform inversion with multicomponent data: from theory to practice: The Leading Edge, **32**, 1040–1054.

Pan, W., Innanen, K. A., and Liao, W., 2016a, Accelerating Hessian-free Gauss-Newton full-waveform inversion via improved preconditioning strategies: SEG Expanded Abstracts, 1455–1461.

Pan, W., Innanen, K. A., and Margrave, G. F., 2014, A comparison of different scaling methods for least-squares migration/inversion: EAGE Expanded Abstracts, We G103 14.

Pan, W., Innanen, K. A., Margrave, G. F., Fhler, M. C., Fang, X., and Li, J., 2015, Estimation of elastic constants in HTI media using Gauss-Newton and Full-Newton multi-parameter full waveform inversion: SEG Technical Program Expanded Abstracts, 1177–1182.

Pan, W., Innanen, K. A., Margrave, G. F., Fhler, M. C., Fang, X., and Li, J., 2016b, Estimation of elastic constants for HTI media using Gauss-Newton and full-Newton multiparameter full-waveform inversion: Geophysics, **81**, R275–R291.

Pan, W., Innanen, K. A., Margrave, G. F., and Keating, S., 2016c, Mitigate cycle-skipping for full-waveform inversion by band-limited impedance inversion and POCS: GeoConvention, 1–5.

Plessix, R. E., and Mulder, W. A., 2004, Frequency-domain finite-difference amplitude-preserving migration: Geophysical Journal International, **157**, 975–987.

Pratt, R. G., Shin, C., and Hicks, G. J., 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341–362.

Saad, Y., 2003, Iterative methods for sparse linear systems: SIAM.

Sainath, T. N., Horesh, L., Kingsbury, B., Aravkin, A. Y., and Ramabhadran, B., 2013, Accelerating Hessian-free optimization for Deep Neural Networks by implicit preconditioning and sampling: IEEE Workshop on Automatic Speech Recognition and Understanding, 303–308.

Santosa, F., and Symes, W. W., 1988, Computation of the Hessian for least-squares solutions of inverse problems of reflection seismology: Inverse Problems, **4**, 211–233.

Shanno, D. F., 1970, Conditioning of quasi-Newton methods for function minimization: Mathematics of Computation, **24**, 647–656.

Shin, C., Jang, S., and Min, D., 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.

Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded Hessian: Geophysics, **74**, WCA95–WCA107.

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

Valenciano, A., 2008, Imaging by wave-equation inversion: Ph.D. thesis, Stanford University.

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

Wu, S., Wang, Y., Zheng, Y., and Chang, X., 2015, Limited-memory BFGS based least-squares pre-stack Kirchhoff depth migration: Geophysical Journal International, **202**, 738–747.

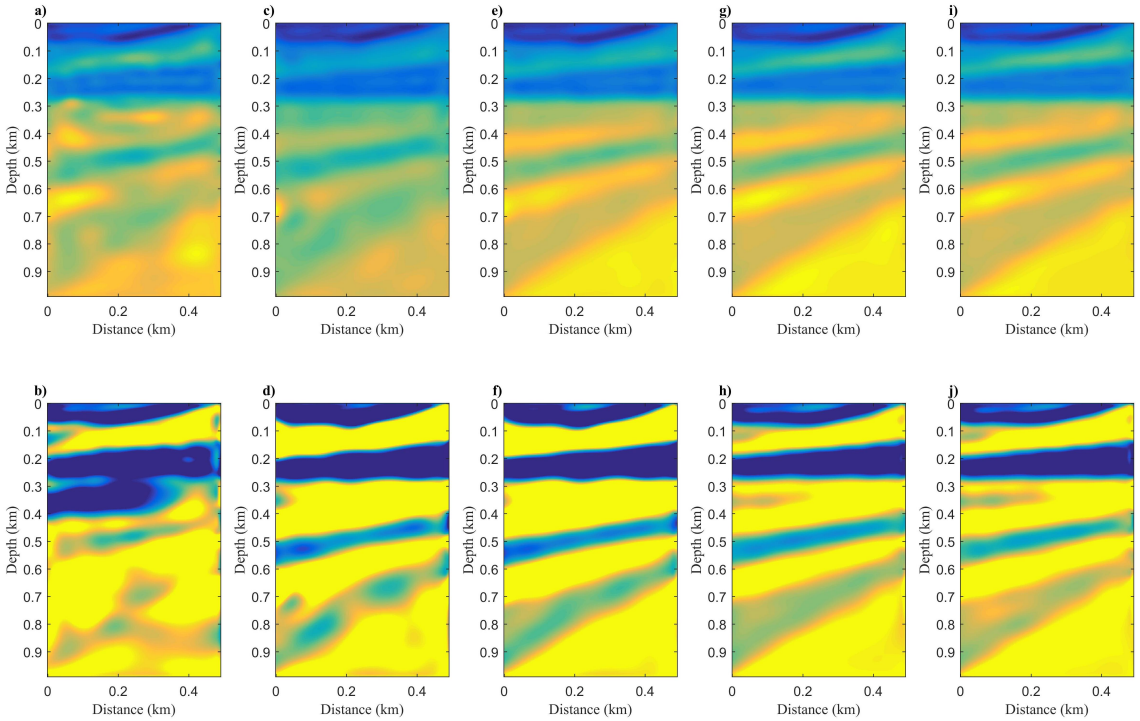

FIG. 11. (a) and (b) show the inverted velocity and density models using SD method; (c) and (d) show the inverted velocity and density models using NCG method; (e) and (f) show the inverted velocity and density models using l-BFGS method. (g) and (h) show the inverted velocity and density models using non-preconditioned HFGN method. (i) and (j) show the inverted velocity and density models using preconditioned HFGN method.