Study of crosstalk reduction in multiparameter acoustic FWI

Qi Hu and Kris Innanen

ABSTRACT

Crosstalk is the phenomenon in which data signatures of different physical properties are confused in full-waveform inversion (FWI). The challenges associated with crosstalk are a major obstacle to the effective implementation of multiparameter FWI. In this study, we focus on acoustic media with variable density, and discuss the reduction of parameter crosstalk from three aspects: optimization method, acquisition geometry, and model parameterization. Optimization methods such as steepest-descent, conjugate gradient and quasi-Newton methods are compared in terms of the inversion quality and computational cost. For geometry, we show how sub-surface sources and receivers provided by vertical seismic profile (VSP) and crosswell help mitigate the crosstalk from surface seismic. Scattering patterns of four parameterizations: velocity-density, impedance-density, impedance-velocity, and bulk modulus-velocity, are illustrated and employed to study their capabilities relative to crosstalk.

INTRODUCTION

With the full wavefield being accounted for in full-waveform inversion (FWI), it is naturally to include more realistic physics in the forward modeling to better match the observed data, such as viscosity, elasticity and anisotropic effects (Operto et al., 2013; Alkhalifah and Plessix, 2014). Multiparameter inversion has become feasible to invert parameter classes other than P-wave velocity, such as density, attenuation, shear velocity. Nevertheless, adding more parameter classes in FWI can increase the ill-posedness of the inverse problem (Virieux and Operto, 2009), and inverting multiparameter is much more complicated than monoparameter inversion due to the potential presence of trade-off/crosstalk between different parameter classes (Geng et al., 2018).

When two parameters have similar scattering patterns, their gradient updates resemble each other, making it difficult to decipher between the two (Innanen, 2014). Newton-based optimization methods, which incorporate the multiparameter inverse Hessian, can alter the update in any one parameter to accommodate the variational properties of all others (Métivier et al., 2013). However, explicit calculation, storage and inversion of Hessian are computationally unaffordable for large-scale inverse problem (Pan et al., 2018). As a low-rank approximation of Hessian, Quasi-Newton methods (e.g., L-BFGS method) approach the inverse Hessian iteratively by storing the changes of the gradient and model from a number of previous iterations (Nocedal, 1980). For multiparameter FWI, the L-BFGS algorithm provides a suitable scaling of the gradients for each parameter class. A comparison between the conjugate-gradient method and the L-BFGS method for a realistic onshore application of multiparameter elastic FWI is shown in Brossier et al. (2009).

Another approach of mitigating parameter crosstalk is to use a suitable model parameterization, where the classes of unknown properties should be as uncorrelated as possible (Keating and Innanen, 2017). Scattering patterns, which represent the analytic solutions

of the Fréchet derivative wavefields due to different physical parameters, have been widely studied for selecting the optimal parameterization (Wu and Aki, 1985; Tarantola, 1986; Forgues and Lambaré, 1997). Tarantola (1986) examines the resolving abilities of various parameterizations for isotropic-elastic FWI. His analysis suggests that the velocity-density parameterization is more appropriate for inversion with large-offset data, whereas the impedance–density parameterization is more suitable for near-offset data. Virieux and Operto (2009) show the radiation pattern of various parameterizations in acoustic FWI. They point out that velocity and density are difficult to recover from short-offset data, and if wide-aperture data are available, velocity and impedance might provide the most judicious parameterization because they scatter energy for different aperture bands.

In this paper, we first review the forward and inverse problems associated with FWI in acoustic media with variable density. Optimization methods such as gradient- and Newton-based methods are explained. The expressions of sensitivity kernels for various parameterizations are derived. We then carry out synthetic experiments to compare different optimizations methods, acquisition geometries, and model parameterizations in terms of their capability relative to crosstalk in multiparameter inversion.

THEORIES AND METHODS

Frequency-domain forward modeling

We use the frequency domain acoustic wave equation to describe wave motions (Marfurt, 1984):

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

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, $K=\rho v^2$ is bulk modulus, ρ is density, v is velocity, $\mathbf{u}(\mathbf{x},\mathbf{x}_s,\omega)$ denotes the pressure wavefield at position $\mathbf{x}$, and $f_s(\omega)\delta(\mathbf{x}-\mathbf{x}_s)$ means the source signature at position $\mathbf{x}_s$. Rewrite the model parameters with different classes into a vector $\mathbf{m}$, the discretized wave equation can be written in a matrix form:

$$\mathbf{A}(\mathbf{m},\omega)\mathbf{u}(\mathbf{x},\mathbf{x}_s,\omega)=f(\mathbf{x}_s,\omega), \tag{2}$$

where $\mathbf{A}(\mathbf{m},\omega)$ is the impedance matrix. As the number of non-zero diagonals is related to the finite-difference scheme, e.g., in a five-point finite difference scheme, the impedance matrix has five non-zero diagonals. The wavefield can be obtained by the inverse of the impedance matrix, which is usually replaced by the direct matrix factorization methods, such as the LU decomposition.

Theories of full-waveform inversion: Gradient and Hessian

As a non-linear least-squares optimization problem, FWI seeks to estimate subsurface parameters through an iterative process by minimizing the difference between the synthetic

data $\mathbf{d}_{syn}$ and observed data $\mathbf{d}_{obs}$. The misfit function is formulated in a least-squares form:

$$\Phi(\mathbf{m}) = \frac{1}{2}\sum_{ns}\sum_{n\omega}\|\mathbf{d}_{obs}(\mathbf{x}_s,\omega) - \mathbf{d}_{syn}(\mathbf{m},\mathbf{x}_s,\omega)\|^2 = \frac{1}{2}\delta\mathbf{d}^T\delta\mathbf{d}^*, \tag{3}$$

where $\mathbf{d}_{syn}(\mathbf{m},\mathbf{x}_s,\omega) = \mathbf{R}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$ is the synthetic data generated using the current model $\mathbf{m}$, $\mathbf{R}$ is the sampling operator that extracts the wavefields at receivers' locations, ^{T} is the transpose operator and * is the conjugate operator.

The Newton optimization approach is based on the second order Taylor-Lagrange expansion of the misfit function:

$$\Phi(\mathbf{m}+\triangle\mathbf{m}) \approx \Phi(\mathbf{m}) + \mathbf{g}^T\triangle\mathbf{m} + \frac{1}{2}\triangle\mathbf{m}^T\mathbf{H}\triangle\mathbf{m}, \tag{4}$$

where $\triangle\mathbf{m}$ is the search direction, $\mathbf{g}$ and $\mathbf{H}$ indicate the gradient and Hessian. To minimize the quadratic approximation of the misfit function, the updated model at $(k+1)$th iteration can be written as

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \mu_k\triangle\mathbf{m}_k, \tag{5}$$

where μ_k is the step length, a scalar constant calculated by a line-search method. The search direction $\triangle\mathbf{m}_k$ is the solution of the Newton linear system:

$$\mathbf{H}_k\triangle\mathbf{m}_k = -\mathbf{g}_k. \tag{6}$$

Here we show how to derive the gradient and Hessian.

The gradient of the misfit function is the first-order partial derivative of the misfit function with respect to model parameter:

$$\mathbf{g} = \frac{\partial\Phi(\mathbf{m})}{\partial\mathbf{m}} = -\Re\left\{\sum_{ns}\sum_{n\omega}\left[\left(\frac{\partial(\mathbf{R}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega))}{\partial\mathbf{m}}\right)^T\delta\mathbf{d}^*\right]\right\}, \tag{7}$$

where $\mathbf{J} = \frac{\partial(\mathbf{R}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega))}{\partial\mathbf{m}}$ is the Fréchet derivative matrix (or sensitivity matrix), and the operator $\Re$ is taken to ensure the gradient remaining real. To calculate $\mathbf{J}$, we take the partial derivative of equation 2 with respect to model parameters:

$$\mathbf{A}(\mathbf{m},\omega)\frac{\partial\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)}{\partial\mathbf{m}} = -\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega). \tag{8}$$

This shows that the first-order partial derivative of the wavefield $\frac{\partial\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)}{\partial\mathbf{m}}$can be obtained by solving the wave equation with a virtual source :

$$f^g = -\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega). \tag{9}$$

The radiation pattern of each parameter class is included in the virtual source, and the calculation for it depends on the details of the finite approximation method. Equation 8

also indicates that the first-order partial derivative of the wavefield with respect to each model position can be interpreted as the wavefield scattered by a small perturbation of the parameter at this position (Geng et al., 2017).

Substituting the virtual source back to the gradient, since only the real part of the complexed-valued vector is taken to obtain the gradient, the gradient can be calculated using the adjoint (conjugate transpose) of the impedance matrix:

$$\mathbf{g} = \Re\left\{\sum_{ns}\sum_{n\omega}\mathbf{u}^{\dagger}(\mathbf{m},\mathbf{x}_s,\omega)\frac{\partial\mathbf{A}^{\dagger}(\mathbf{m},\omega)}{\partial\mathbf{m}}\lambda(\mathbf{m},\omega)\right\}, \tag{10}$$

where $\lambda(\mathbf{m},\omega)$ is the adjoint variable/adjoint state that is the solution of the adjoint equation (Plessix, 2006):

$$\mathbf{A}^{\dagger}(\mathbf{m},\omega)\lambda(\mathbf{m},\omega) = \mathbf{R}^{\dagger}\delta\mathbf{d}, \tag{11}$$

where † stands for the adjoint operator, and $\mathbf{R}^{\dagger} = \mathbf{R}^{T}$ since $\mathbf{R}$ is real defined.

For multiparameter FWI, the Hessian is a large block matrix, with each block representing the second order derivative of the misfit function with respect to parameter classes:

$$\begin{aligned}\mathbf{H}_{\mathbf{m}_1\mathbf{m}_2}(\mathbf{x},\mathbf{x}') = \Re\Bigg\{\sum_{ns}\sum_{n\omega}\Bigg[&\left(\frac{\partial\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)}{\partial\mathbf{m}_1(\mathbf{x})}\right)^{\dagger}\mathbf{R}^{\dagger}\mathbf{R}\frac{\partial\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)}{\partial\mathbf{m}_2(\mathbf{x}')}\\ &-\left(\frac{\partial^2\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)}{\partial\mathbf{m}_1(\mathbf{x})\partial\mathbf{m}_2(\mathbf{x}')}\right)^{\dagger}\mathbf{R}^{\dagger}\delta\mathbf{d}\Bigg]\Bigg\}.\end{aligned} \tag{12}$$

When $\mathbf{m}_1 = \mathbf{m}_2$, it indicates the element in diagonal blocks, and when $\mathbf{m}_1 \neq \mathbf{m}_2$, it indicates the element in off-diagonal blocks. The first-order term $\mathbf{H}^{L}$ measures the correlations of two Fréchet derivative wavefields and is essential in overcoming crosstalk in multiparameter FWI.

Optimization methods

The full Newton method uses the quadratic search direction from equation 6: $\triangle\mathbf{m}_k = -\mathbf{H}_k^{-1}\mathbf{g}_k$. The Gauss-Newton method approximates the full Hessian by only accounting for the first term in equation 12. Although the Newton methods benefit from fast convergence rate, the computation, storage, and inversion of Hessian at each iteration are prohibitively expensive, this limits their applications for large-scale inverse problems (Pan et al., 2018).

Gradient-based methods (e.g., steepest-descent (SD) and conjugate-gradient (CG)) are computationally more attractive when inverting a large number of unknown model parameters (Pan et al., 2017). SD simply determines the search direction to be the negative of gradient. In CG, the search direction is a linear combination of current gradient and the previous search direction. However, Pratt et al. (1998) claim that the gradient method can fail to converge toward an acceptable model, however many iterations. They interpret this as the result of the difficulty of estimating a reliable step length.

As a low-rank approximation of the Hessian, the L-BFGS method is an attractive alternative to the Newton-type and gradient-based methods by approximating the inverse Hessian iteratively instead of constructing the Hessian explicitly (Brossier et al., 2009). Using

a pair of vectors $\mathbf{s}_{k+1} = \mathbf{m}_{k+1} - \mathbf{m}_k$ and $\mathbf{y}_{k+1} = \mathbf{g}_{k+1} - \mathbf{g}_k$ that indicate the model and gradient changes that satisfy the condition $\mathbf{s}_k^\dagger \mathbf{y}_k > 0$, the inverse Hessian approximation $\mathbf{H}_{k+1}^{-1}$ is given by

$$\mathbf{H}_{k+1}^{-1} = \mathbf{v}_k^\dagger \mathbf{H}_k^{-1} \mathbf{v}_k^\dagger + \mathbf{w}_k \mathbf{s}_k \mathbf{s}_k^\dagger, \tag{13}$$

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 model and gradient are stored for a limited number (typically < 10) of previous iterations. For multiparameter FWI, the L-BFGS algorithm provides a suitable scaling of the gradient computed for each parameter class and can help reduce crosstalk.

Sensitivity kernels of different parameterizations

To consider the model perturbation with the form of relative variation, for the bulk modulus-density parameterization, we define the model as

$$s_k(\mathbf{x}) = \ln K(\mathbf{x}), s_\rho(\mathbf{x}) = \ln \rho(\mathbf{x}). \tag{14}$$

So the perturbations of modulus and density can be expressed as

$$\delta s_k = \frac{\delta K}{K}, \delta s_\rho = \frac{\delta \rho}{\rho}. \tag{15}$$

Consider a model perturbation at position $\mathbf{x}$, with the Born approximation, the perturbed pressure wavefield for a source-receiver couple at frequency ω is given by

$$\delta u(\mathbf{x}_g, \mathbf{x}_s, \omega) = \int_{\Omega(\mathbf{x})} (\mathcal{K}_{K\rho-K}(\mathbf{x}) \delta s_k(\mathbf{x}) + \mathcal{K}_{K\rho-\rho}(\mathbf{x}) \delta s_\rho(\mathbf{x})) d\mathbf{x}, \tag{16}$$

where $\mathcal{K}_{K\rho-K}$ and $\mathcal{K}_{K\rho-\rho}$ represent the sensitivity kernels with respect to the bulk modulus K and density ρ, and can be expressed as

$$\begin{aligned} \mathcal{K}_{K\rho-K}(\mathbf{x}) &= -\frac{\omega^2}{K} G(\mathbf{x}_g, \mathbf{x}, \omega) G(\mathbf{x}, \mathbf{x}_s, \omega), \\ \mathcal{K}_{K\rho-\rho}(\mathbf{x}) &= \frac{1}{\rho} \nabla G(\mathbf{x}_g, \mathbf{x}, \omega) \cdot \nabla G(\mathbf{x}, \mathbf{x}_s, \omega), \end{aligned} \tag{17}$$

where $G(\mathbf{x}, \mathbf{x}_s, \omega)$ and $G(\mathbf{x}_g, \mathbf{x}, \omega)$ denote the source-side and receiver-side Green's functions.

For the velocity-density parameterization, we define

$$s_v(\mathbf{x}) = \ln v(\mathbf{x}), s_\rho(\mathbf{x}) = \ln \rho(\mathbf{x}). \tag{18}$$

Using the chain rule:

$$\frac{\delta K}{K} \approx 2\frac{\delta v}{v} + \frac{\delta \rho}{\rho}, \tag{19}$$

the corresponding sensitivity kernels are:

$$\begin{aligned}\mathcal{K}_{v\rho-v}(\mathbf{x}) &= -\frac{2\omega^2}{\rho v^2}G(\mathbf{x}_g,\mathbf{x},\omega)G(\mathbf{x},\mathbf{x}_s,\omega),\\ \mathcal{K}_{v\rho-\rho}(\mathbf{x}) &= -\frac{\omega^2}{\rho v^2}G(\mathbf{x}_g,\mathbf{x},\omega)G(\mathbf{x},\mathbf{x}_s,\omega)+\frac{1}{\rho}\nabla G(\mathbf{x}_g,\mathbf{x},\omega)\cdot\nabla G(\mathbf{x},\mathbf{x}_s,\omega).\end{aligned} \tag{20}$$

Likewise, sensitivity kernels of the other four parameterizations for acoustic media can be derived.

Impedance-density:

$$\begin{aligned}\mathcal{K}_{I_{\mathrm{p}}\rho-I_{\mathrm{p}}}(\mathbf{x}) &= -\frac{2\omega^2\rho}{I_{\mathrm{p}}^2}G(\mathbf{x}_g,\mathbf{x},\omega)G(\mathbf{x},\mathbf{x}_s,\omega),\\ \mathcal{K}_{v\rho-\rho}(\mathbf{x}) &= \frac{\omega^2\rho}{I_{\mathrm{p}}^2}G(\mathbf{x}_g,\mathbf{x},\omega)G(\mathbf{x},\mathbf{x}_s,\omega)+\frac{1}{\rho}\nabla G(\mathbf{x}_g,\mathbf{x},\omega)\cdot\nabla G(\mathbf{x},\mathbf{x}_s,\omega).\end{aligned} \tag{21}$$

Impedance-velocity:

$$\begin{aligned}\mathcal{K}_{I_{\mathrm{p}}v-I_{\mathrm{p}}}(\mathbf{x}) &= -\frac{\omega^2}{I_{\mathrm{p}}\cdot v}G(\mathbf{x}_g,\mathbf{x},\omega)G(\mathbf{x},\mathbf{x}_s,\omega)-\frac{v}{I_{\mathrm{p}}}\nabla G(\mathbf{x}_g,\mathbf{x},\omega)\cdot\nabla G(\mathbf{x},\mathbf{x}_s,\omega),\\ \mathcal{K}_{I_{\mathrm{p}}v-v}(\mathbf{x}) &= -\frac{\omega^2}{I_{\mathrm{p}}\cdot v}G(\mathbf{x}_g,\mathbf{x},\omega)G(\mathbf{x},\mathbf{x}_s,\omega)-\frac{v}{I_{\mathrm{p}}}\nabla G(\mathbf{x}_g,\mathbf{x},\omega)\cdot\nabla G(\mathbf{x},\mathbf{x}_s,\omega).\end{aligned} \tag{22}$$

Modulus-velocity:

$$\begin{aligned}\mathcal{K}_{Kv-K}(\mathbf{x}) &= -\frac{\omega^2}{K}G(\mathbf{x}_g,\mathbf{x},\omega)G(\mathbf{x},\mathbf{x}_s,\omega)+\frac{v^2}{K}\nabla G(\mathbf{x}_g,\mathbf{x},\omega)\cdot\nabla G(\mathbf{x},\mathbf{x}_s,\omega),\\ \mathcal{K}_{Kv-v}(\mathbf{x}) &= -\frac{2v^2}{K}\nabla G(\mathbf{x}_g,\mathbf{x},\omega)\cdot\nabla G(\mathbf{x},\mathbf{x}_s,\omega).\end{aligned} \tag{23}$$

Modulus-impedance:

$$\begin{aligned}\mathcal{K}_{KI_{\mathrm{p}}-K}(\mathbf{x}) &= -\frac{\omega^2}{K}G(\mathbf{x}_g,\mathbf{x},\omega)G(\mathbf{x},\mathbf{x}_s,\omega)-\frac{K}{I_{\mathrm{p}}^2}\nabla G(\mathbf{x}_g,\mathbf{x},\omega)\cdot\nabla G(\mathbf{x},\mathbf{x}_s,\omega),\\ \mathcal{K}_{KI_p-I_{\mathrm{p}}}(\mathbf{x}) &= -\frac{2K}{I_{\mathrm{p}}^2}\nabla G(\mathbf{x}_g,\mathbf{x},\omega)\cdot\nabla G(\mathbf{x},\mathbf{x}_s,\omega).\end{aligned} \tag{24}$$

The spatial derivative term $\nabla G(\mathbf{x}_g,\mathbf{x},\omega)\cdot\nabla G(\mathbf{x},\mathbf{x}_s,\omega)$ satisfies

$$\nabla G(\mathbf{x}_g,\mathbf{x},\omega)\cdot\nabla G(\mathbf{x},\mathbf{x}_s,\omega) = -\cos\theta\cdot\frac{\omega^2}{v^2}G(\mathbf{x}_g,\mathbf{x},\omega)G(\mathbf{x},\mathbf{x}_s,\omega), \tag{25}$$

where θ is the source-receiver opening angle.

Substitute equation 25 into the above sensitivity kernels, we obtain the radiation patterns for the six parameterizations (He et al., 2018):

$$
\begin{aligned}
Q_{v-\rho} &= [-2, -2\cos^2(\theta/2)], \\
Q_{K-\rho} &= [-1, \cos\theta], \\
Q_{I_\mathrm{p}-\rho} &= [-2, -2\sin^2(\theta/2)], \\
Q_{I_\mathrm{p}-v} &= [-2\cos^2(\theta/2), -2\sin^2(\theta/2)], \\
Q_{K-v} &= [-2\cos^2(\theta/2), -2\cos\theta], \\
Q_{K-I_\mathrm{p}} &= [-2\sin^2(\theta/2), -2\cos\theta].
\end{aligned}
\tag{26}
$$

NUMERICAL EXAMPLES

Most numerical tests in this section are performed on the Gaussian-anomaly model (Figure 1), where we have a velocity anomaly on the left and a density anomaly on the right. The initial velocity and density models are homogeneous, with a constant velocity of 3000 m/s and a constant density of 2000 $\mathrm{kg/m^3}$. The model is constructed of 50 $\times$ 100 grid cells with a grid interval of 10 m in horizontal and vertical directions. A total of 10 sources are deployed from 20 to 980 m with an interval of 100 m and a depth of 20 m. A total of 100 receivers are distributed on the surface from 10 to 1000 m with an interval of 10 m and a depth of 20 m. 13 frequencies from 3 HZ to 15 HZ with 20 iterations per frequency are used. The defaulted optimization method, acquisition geometry and parameterization are the L-BFGS method, surface-only sources and receivers, and velocity-density parameterization, respectively.

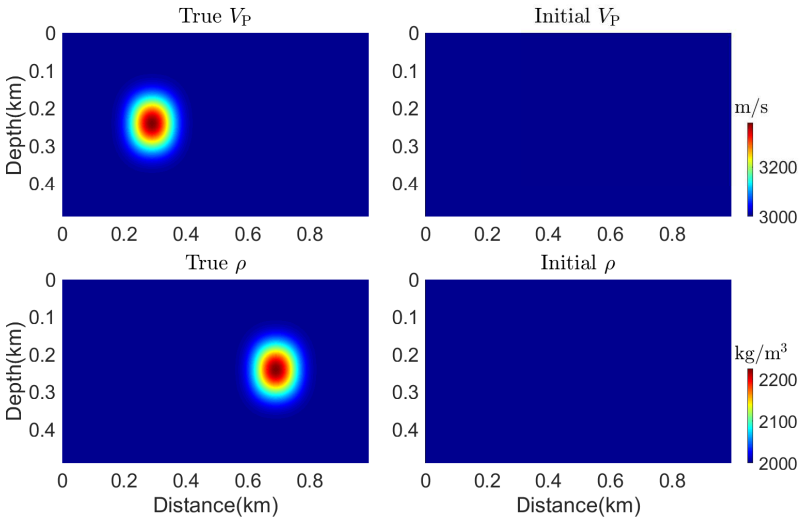

FIG. 1. True and initial velocity and density models

Optimization Methods

Figure 2 shows the inversion results with the steepest descent (SD), conjugate gradient (CG) and L-BFGS methods. The stopping criteria is set to be a maximum iteration of 200.

We observe that the mappings between velocity and density appear in all figures. Compared with velocity, density is more difficult to recover using any of the three methods. Due to a stronger mapping from density to velocity and a dimmer reconstructed density, SD is considered to be less efficient than CG and L-BFGS.

Figure 3 shows the convergence history of the three methods. The uphill steps in data residuals imply a transition from one frequency band to the next. The normalized model errors are calculated by dividing the errors of the estimated model by the errors of the initial model. We observe that L-BFGS converges slightly faster than CG and much faster than SD. Besides, the computation time in this case is 397 s for SD, 508 s for CG, and 308 s for L-BFGS, this illustrates that L-BFGS outperforms SD and CG on both accuracy and computational cost.

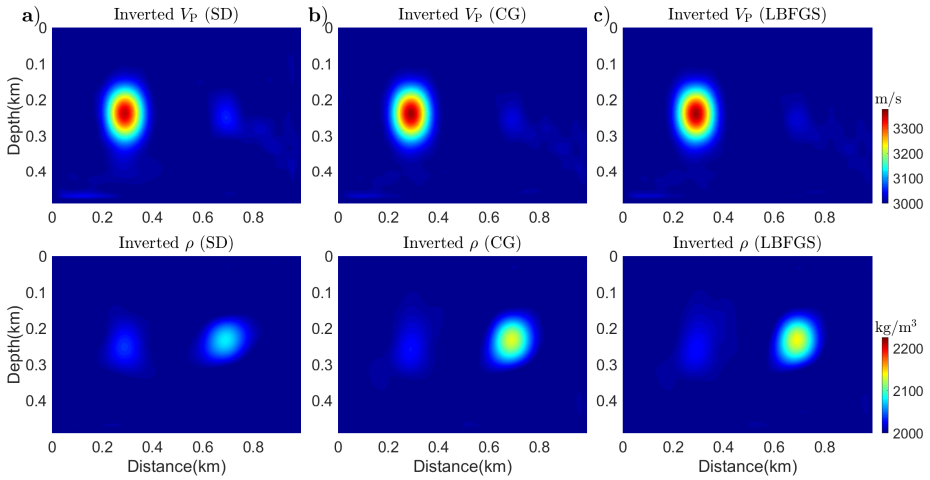

FIG. 2. Comparison of inversion results by different optimization methods: a) SD, b) CG, and c) L-BFGS.

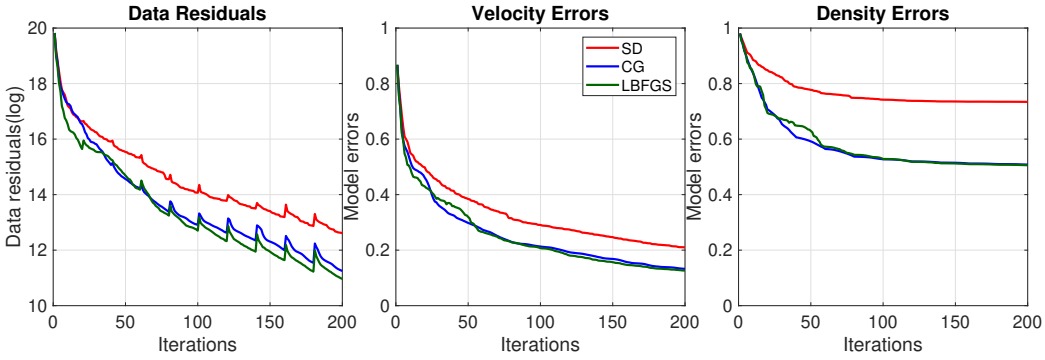

FIG. 3. Comparison of the convergence history for different optimization methods.

The three optimization methods are tested on another model, as shown in Figure 4. We have a layered velocity model and a Gauss-anomaly density model. The initial velocity is obtained by smoothing the true velocity model and the initial density model is set to be homogeneous. Figures 5 and 6 show the inversion results and convergence histories. Consistent with the results in previous example, the three methods have a similar reduction

trend of data misfit and velocity errors, but L-BFGS takes a greater advantage of recovering density. Given the computation time of this example is 284 s for SD, 447 s for CG and 228 s for L-BFGS, we conclude that L-BFGS is not only faster but also more effective for reducing crosstalk.

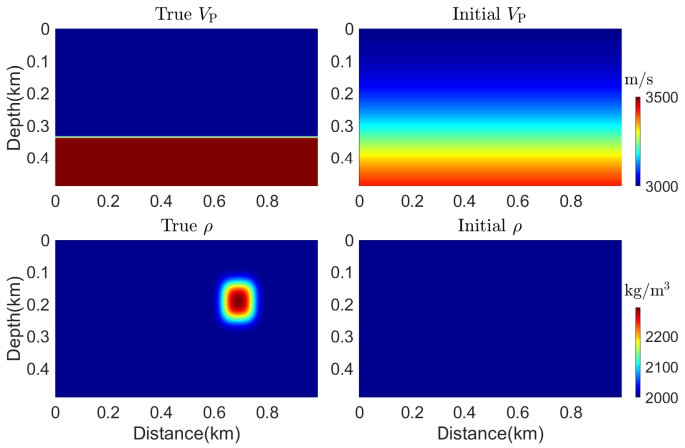

FIG. 4. True and initial models of velocity and density

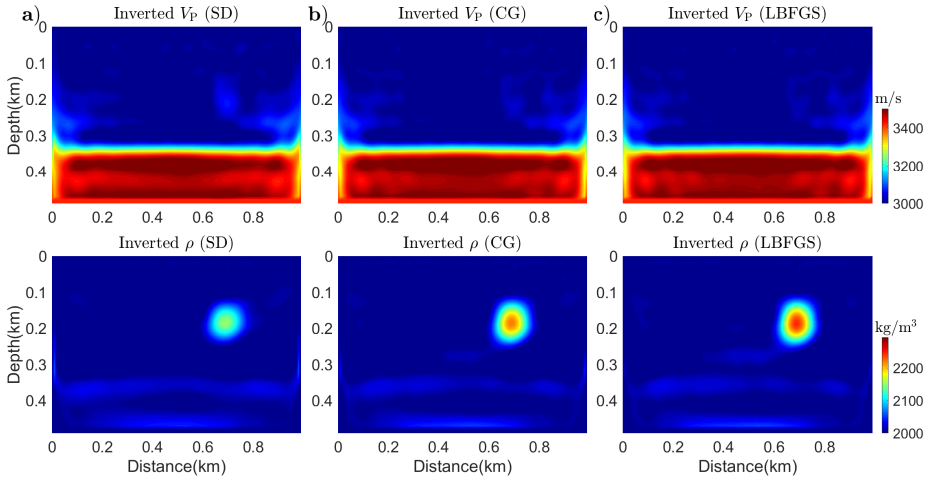

FIG. 5. Comparison of inversion results by different optimization methods: a) SD, b) CG, and c) L-BFGS.

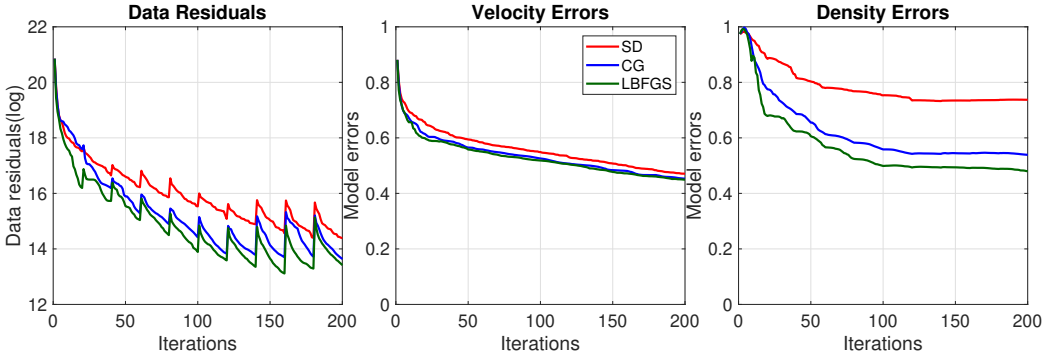

FIG. 6. Comparison of the convergence history for different optimization methods.

Acquisition Geometries

The importance of acquisition geometry to the reduction of parameter crosstalk in FWI can be illustrated by radiation patterns. Figure 7 shows the radiation patterns of velocity and density. We observe that the perturbation of velocity scatters the same energy for different angles, and density has the same radiation pattern as velocity at short apertures but does not scatter energy at wide apertures. This implies that the two parameters are difficult to reconstruct using only short-offset data.

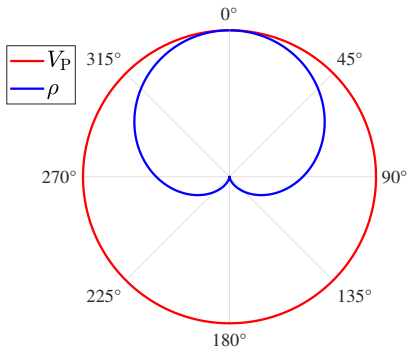

FIG. 7. Radiation pattern of the velocity-density parameterization.

Here we design different acquisition geometries. First, two theoretical configurations are used to compare with the surface seismic. As illustrated in Figure 8, the gray frame denotes the model space, the red and blue lines denote the distribution of sources and receivers, and the two spheres denote the velocity and density anomalies. In Figure 8a, we have sources and receivers distributed only on the surface. In Figure 8b, receivers are spread to cover all four boundaries of the model space. In Figure 8c, sources and receivers are both extended to cover all boundaries. Figure 9 shows the inverted velocity and density under the corresponding acquisition geometry in Figure 8.

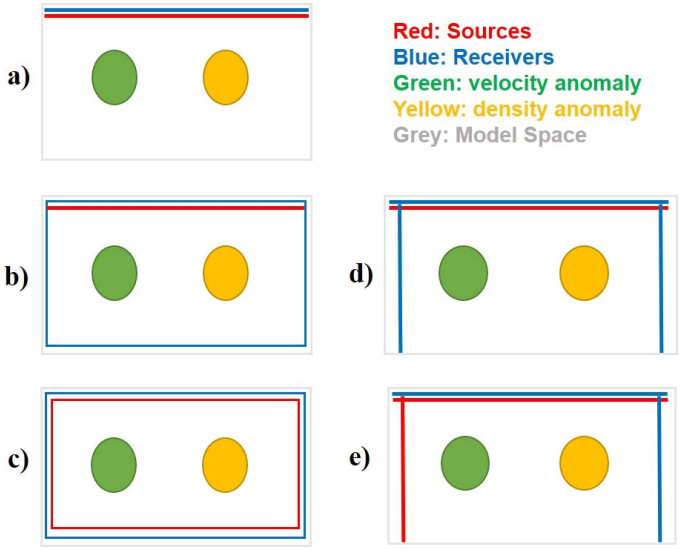

FIG. 8. Acquisition geometries.

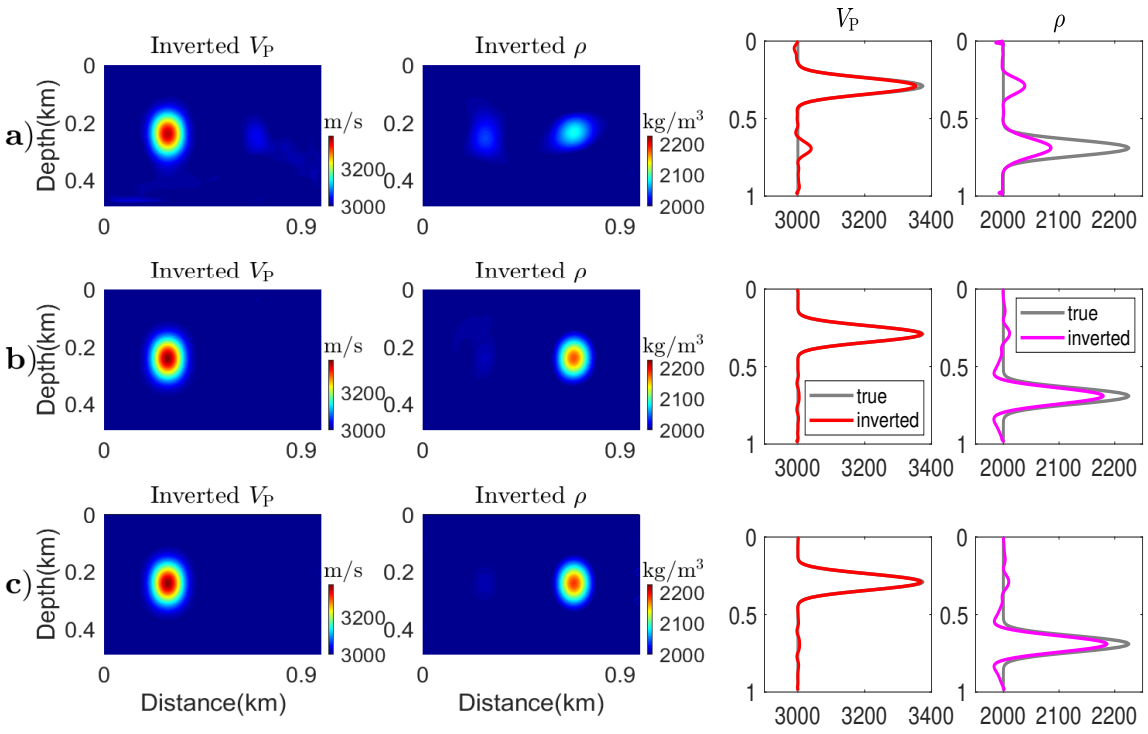

FIG. 9. Comparison of the inverted velocity and density using different acquisition geometries.

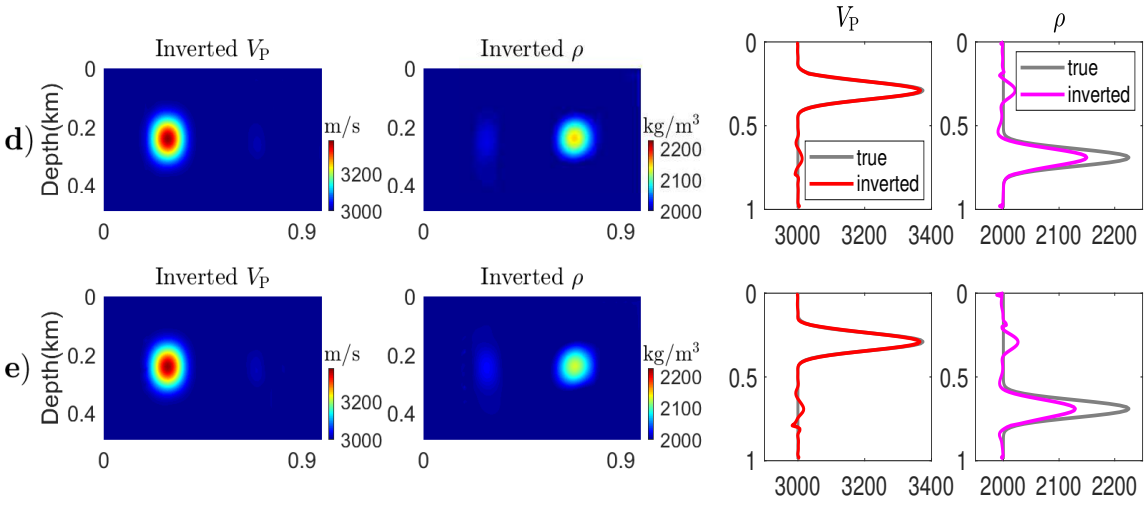

FIG. 10. Comparison of the inverted velocity and density using d) VSP + surface seismic and e) crosswell + surface seismic.

In case a), we see that both the inverted velocity and density suffer from crosstalk. In case b), the inverted models have been largely enhanced because the mapping from density to velocity becomes negligible and the mapping from velocity to density is suppressed. In case c), the inverted models are slightly more accurate than the results in case b). We extract the results along the depth of 250 m, as shown in the right side of Figure 9. An almost perfect match between the inverted velocity and the true model is found in case b) and c). The improvement is significant comparing b) with a), but is minor comparing c) with b). This is because we add transmission data to case b), but a transition from b) to c) won't provide data of extra scattering angles.

Although the geometries illustrated in Figures 8b and 8c help mitigate the crosstalk from surface seismic, they are not feasible in reality. A good substitution would be to add sub-surface sources and receivers through wells. Techniques such as the vertical seismic profile (VSP) and crosswell are combined with surface seismic in Figures 8d and 8e. In VSP we have two receiving wells near the left and right edges of the model. In crosswell we use one source well on the left and one receiving well on the right. The inversion results (Figure 10) are largely improved compared to the results in surface seismic. Figure 11 shows the comparison of the inverted velocity and density using surface seismic and VSP + surface seismic for the model in Figure 4. Obvious crosstalk can be found in Figure 11a, but they are highly suppressed in Figure 11b.

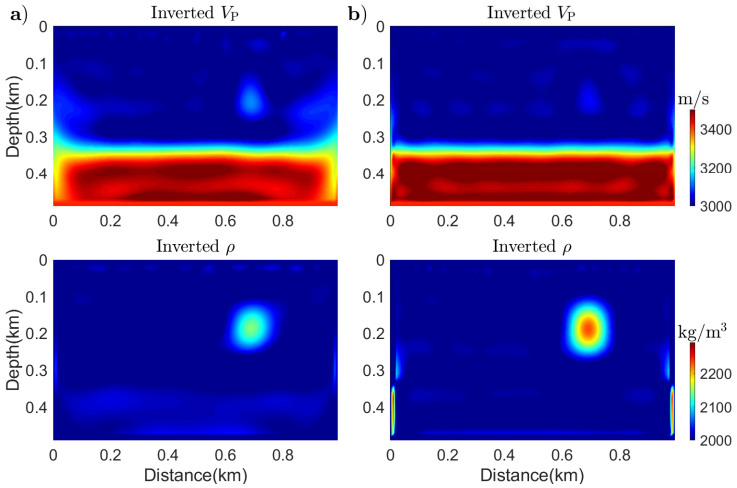

FIG. 11. Comparison of the inverted velocity and density using a) surface seismic and b) VSP + surface seismic.

Parameterizations

Here we consider 4 parameterizations: velocity-density ($V_\mathrm{P} - \rho$), impedance-density ($I_\mathrm{P} - \rho$), impedance-velocity ($I_\mathrm{P} - V_\mathrm{P}$), and bulk modulus-velocity ($K - V_\mathrm{P}$). Their radiation patterns (Figure 12) are calculated based on equation 26. Contrary to $V_\mathrm{P} - \rho$, in $I_\mathrm{P} - \rho$ density has the same pattern as impedance at large offsets, but scatters minor energies at short apertures. $I_\mathrm{P} - V_\mathrm{P}$ might be the most effective parameterization because the two parameters have opposite scattering patterns. $K - V_\mathrm{P}$ seems to be least favorable because in general the aperture is restricted from being wide enough, and the two parameters have similar patterns at short apertures.

We test these parameterizations on the model in Figure 13. In each parameterization, two parameters would be inverted directly, the other two are calculated afterwards. The acquisition geometry of surface-only sources and receivers is adopted. We first compare three parameterizations: $V_\mathrm{P} - \rho$, $I_\mathrm{P} - V_\mathrm{P}$, and $K - V_\mathrm{P}$. As shown in Figure 14, $K - V_\mathrm{P}$ does result in the worst images because the reconstructed anomalies are more distorted and the background are more noisy. The observation is confirmed by extracting the results along the depth of 250 m (Figure 15). We see that all parameters are better estimated using $I_\mathrm{P} - V_\mathrm{P}$ than using $K - V_\mathrm{P}$.

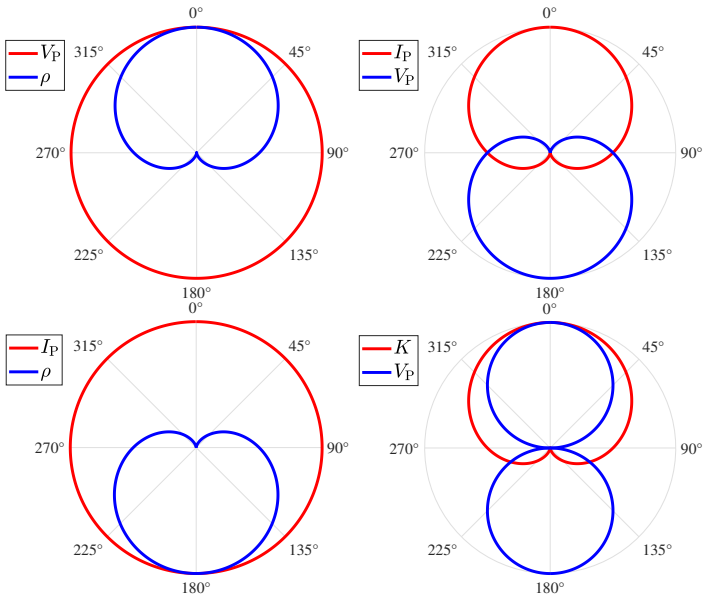

FIG. 12. Radiation patterns of different parameterizations.

To compare the parameterizations quantitatively, we calculate the normalized model errors in each parameterization and rank them, as shown in Figure 16. We note that $K - V_P$ generates the largest errors, and $I_P - V_P$ gives the smallest errors for three parameters: density, impedance, and bulk modulus. Although $V_P - \rho$ is slightly less accurate than $I_P - V_P$ when considering the overall performance, it leads to a better velocity result. We conclude that $I_P - V_P$ and $V_P - \rho$ are the most effective parameterizations in acoustic FWI, the preference of one over another depends on the specific parameter we tend to estimate.

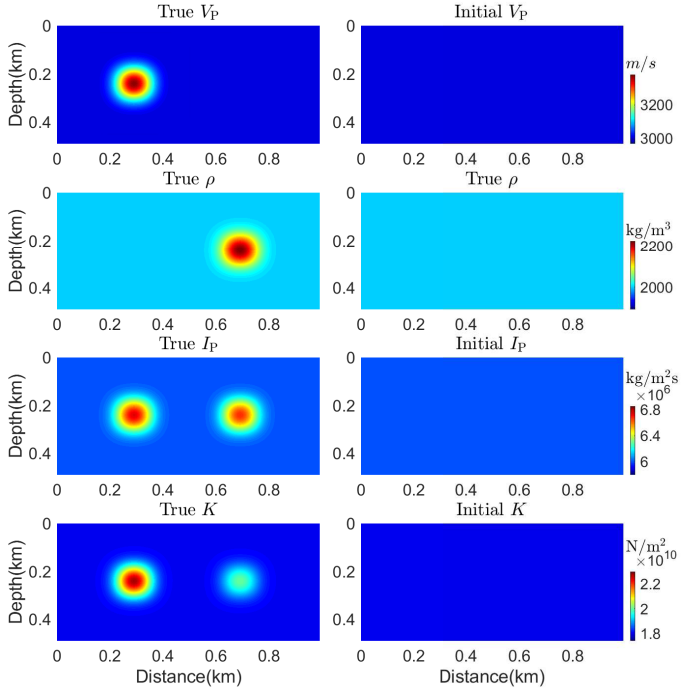

FIG. 13. True and initial models of velocity, density, impedance, and bulk modulus.

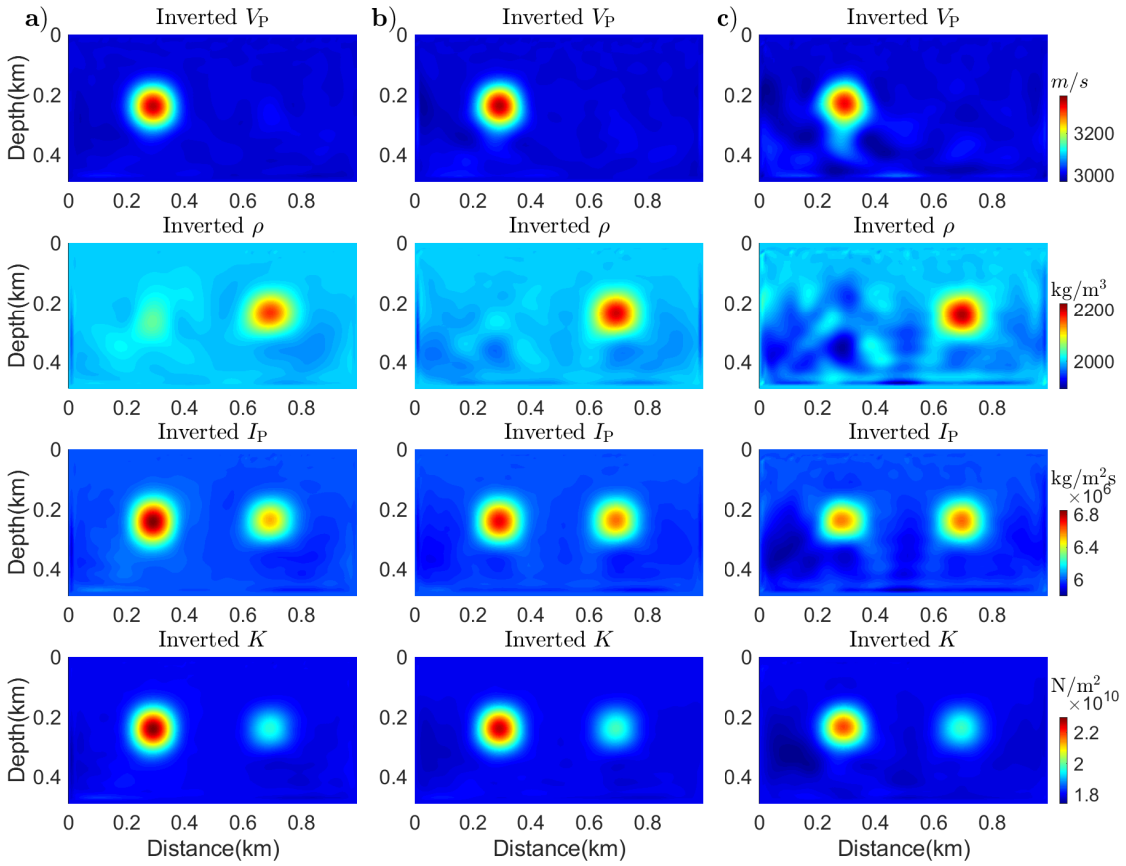

FIG. 14. Inversion results with different parameterizations: a) $V_P - \rho$, b) $I_P - V_P$, c) $K - V_P$.

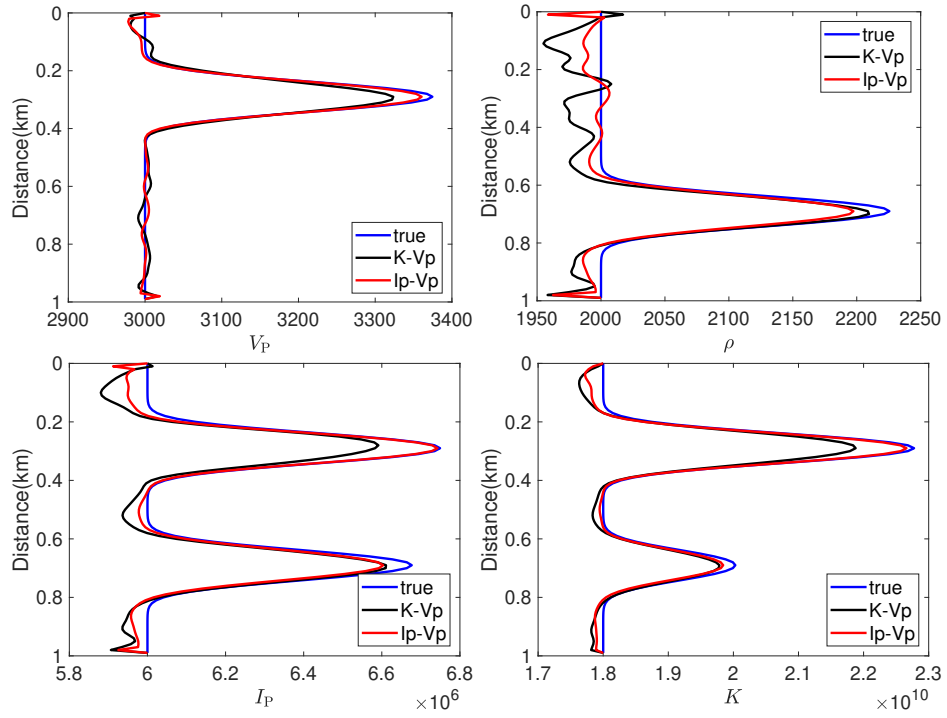

FIG. 15. Inversion results with the $K - V_P$ (blue) and $I_P - V_P$ (red) parameterizations.

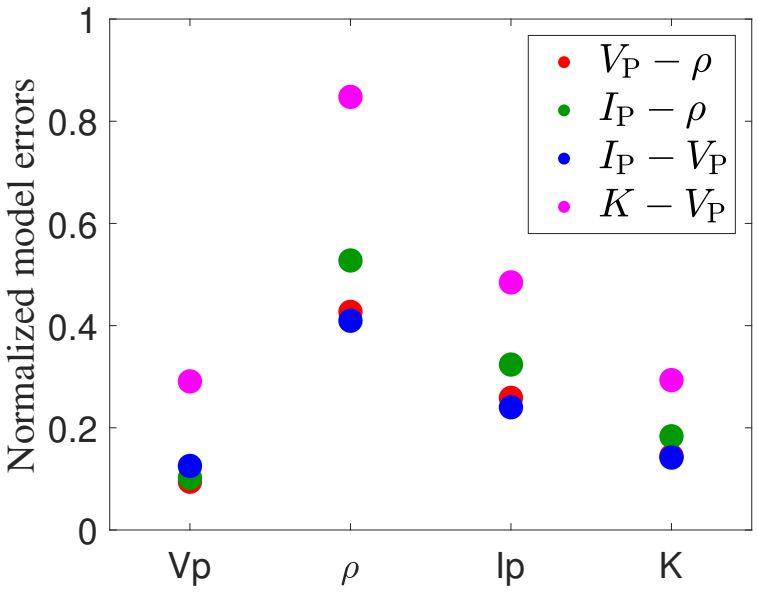

FIG. 16. Model errors in different parameterizations.

CONCLUSIONS

In this paper, we studied three factors that are crucial for reducing crosstalk in multiparameter acoustic FWI: optimization method, acquisition geometry, and parameterization. We demonstrate that the L-BFGS method outperforms SD and CG on both accuracy and computation cost. Based on the analysis of radiation patterns, we show that incorporating surface seismic with sub-surface sources and receivers through VSP and cross-wells help suppress crosstalk. Different parameterizations are compared. For our experiments, $I_P - V_P$ and $V_P - \rho$ are the two most effective parameterizations.

ACKNOWLEDGMENTS

We 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.

REFERENCES

Alkhalifah, T., and Plessix, R.-É., 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study: Geophysics, **79**, No. 3, R91–R101.

Brossier, R., Operto, S., and Virieux, J., 2009, Seismic imaging of complex onshore structures by 2d elastic frequency-domain full-waveform inversion: Geophysics, **74**, No. 6, WCC105–WCC118.

Forgues, E., and Lambaré, G., 1997, Parameterization study for acoustic and elastic ray plus born inversion: Journal of Seismic Exploration, **6**, No. 2-3, 253–277.

Geng, Y., Innanen, K. A., and Pan, W. Y., 2017, Nonlinear multiparameter full waveform inversion based on truncated newton method: CREWES Research Report, **29**, 1–31.

Geng, Y., Pan, W., and Innanen, K. A., 2018, Frequency-domain full-waveform inversion with non-linear descent directions: Geophysical Journal International, **213**, No. 2, 739–756.

He, B., Fang, W., Hu, G., Liu, D., and Sun, S., 2018, Parameterization of acoustic wave equation and strategy for multi-parameter full waveform inversion: Geophysical Prospecting for Petroleum, **57**, No. 5, 705–716.

Innanen, K. A., 2014, Seismic avo and the inverse hessian in precritical reflection full waveform inversion: Geophysical Journal International, **199**, No. 2, 717–734.

Keating, S., and Innanen, K. A., 2017, Crosstalk and frequency bands in truncated newton an-acoustic full-waveform inversion, *in* SEG Technical Program Expanded Abstracts 2017, Society of Exploration Geophysicists, 1416–1421.

Marfurt, K. J., 1984, Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations: Geophysics, **49**, No. 5, 533–549.

Métivier, L., Brossier, R., Virieux, J., and Operto, S., 2013, Full waveform inversion and the truncated newton method: SIAM Journal on Scientific Computing, **35**, No. 2, B401–B437.

Nocedal, J., 1980, Updating quasi-newton matrices with limited storage: Mathematics of computation, **35**, No. 151, 773–782.

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**, No. 9, 1040–1054.

Pan, W., Innanen, K. A., and Geng, Y., 2018, Elastic full-waveform inversion and parametrization analysis applied to walk-away vertical seismic profile data for unconventional (heavy oil) reservoir characterization: Geophysical Journal International, **213**, No. 3, 1934–1968.

Pan, W., Innanen, K. A., and Liao, W., 2017, Accelerating hessian-free gauss-newton full-waveform inversion via l-bfgs preconditioned conjugate-gradient algorithm: Geophysics, **82**, No. 2, R49–R64.

Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, No. 2, 495–503.

Pratt, R. G., Shin, C., and Hick, G., 1998, Gauss–newton and full newton methods in frequency–space seismic waveform inversion: Geophysical Journal International, **133**, No. 2, 341–362.

Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, No. 10, 1893–1903.

Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, No. 6, WCC1–WCC26.

Wu, R.-S., and Aki, K., 1985, Scattering characteristics of elastic waves by an elastic heterogeneity: Geophysics, **50**, No. 4, 582–595.