Interparameter tradeoffs quantification and reduction in isotropic-elastic FWI: synthetic experiments and real dataset application

Wenyong Pan, Yu Geng and Kris Innanen

ABSTRACT

This report aims at evaluating and reducing the interparameter tradeoffs in isotropic-elastic FWI with multiparameter Hessian matrix-vector products. It is revealed that products of multiparameter Hessian off-diagonal blocks with model perturbation vectors, defined as *interparameter contamination kernels*, mainly account for interparameter tradeoff. The multiparameter Hessian is applied to various vectors designed to provide information regarding the strengths and characteristics of interparameter contaminations locally or within the whole volume. Based on these findings, a novel strategy is developed to mitigate the influence of interparameter tradeoffs with approximate contamination kernels. Furthermore, I propose to quantify resolution of the inverted models with approximate eigenvalue volume and extended multiparameter point spread functions (EMPSFs) by preconditioned conjugate-gradient algorithm. Finally, the proposed inversion strategies are applied to invert isotropic-elastic parameters with synthetic data and Hussar practical seismic dataset. Resolution of the inverted models are also evaluated.

INTRODUCTION

Elastic parameters are important for reservoir characterization. Simultaneously reconstructing multiple physical parameters suffers from interparameter tradeoffs arising from the inherent ambiguities among these parameters, which increases the nonlinearity and uncertainty of the inverse problems significantly (Tarantola, 1986; Köhn et al., 2012; Innanen, 2013; Operto et al., 2013; Alkhalifa and Plessix, 2014; Innanen, 2014). This chapter aims at: (1) creating more complete tools for quantifying the interparameter tradeoffs (or parameter crosstalk) than currently exist; (2) evaluating the strengths and characteristics of the interparameter contaminations in isotropic-elastic FWI by applying the multiparameter Hessian to various types of test vectors; (3) developing an effective way to reduce the influence of interparameter tradeoffs based on approximate contamination kernels; (4) quantifying local spatial and interparameter tradeoff of the inverted models with extended multiparameter point spread functions (EMPSFs).

Researchers have devoted intensive efforts to the study of parameter resolution based on analytic solutions of Fréchet derivative wavefields (“scattering" or “radiation" patterns) for different parameter classes (Tarantola, 1986; Gholami et al., 2013b; Alkhalifa and Plessix, 2014; Kamath and Tsvankin, 2014; Podgornova et al., 2015; Oh and Alkhalifah, 2016). Coupling effects appear between two different physical parameters, if the scattered wavefields due to the model perturbations overlap at certain range of scattering angles (Tarantola, 1986). A high-resolution parameterization should have scattering patterns as different as possible (Tarantola, 1986). Gholami et al. (2013a) investigated the scattering patterns of parameters resulting from various parameterizations of multiparameter acoustic FWI. Alkhalifa and Plessix (2014) emphasized the power of horizontal P-wave velocity in reducing the number of parameters for VTI FWI.

Amplitude variations of scattering patterns provide invaluable information for understanding the interparameter coupling effects but also ignore some important aspects due to a series of assumptions including incident plane-wave, homogeneous and isotropic-elastic background, high-frequency approximation, etc (Podgornova et al., 2015). These assumptions are regularly violated in seismic data sets, i.e., finite-frequency effects and traveltime information are not negligible in their influence on parameter resolution; heterogeneities should be considered; spatial correlations of different physical parameters are neglected (Alkhalifa and Plessix, 2014). Overlapping the scattering patterns due to different physical parameters in fact represents only an asymptotic approximation of the crosstalk quantification intrinsic to the Gauss-Newton Hessian (Operto et al., 2013). These limitations of the scattering patterns may result in misunderstandings concerning the interparameter tradeoffs. The problem of isotropic-elastic FWI has been investigated by many researchers (Mora, 1987; Brossier et al., 2009; Köhn et al., 2012; Yuan and Simons, 2014; Borisov and Singh, 2015; Raknes and Arntsen, 2015; Modrak et al., 2016; Pan and Innanen, 2016a,c,b), but many challenges and open questions remain. Density structures are still poorly constrained, which may be caused by the weak sensitivity of traveltime to density variations and strong contaminations from velocity parameters. Some issues associated with the interparameter tradeoffs of isotropic-elastic parameters are actually not explained completely and clearly. Further unanswered questions include:

1. how do the interparameter tradeoffs affect the inversion process ?
2. how to evaluate the strengths and characteristics of the interparameter contaminations quantitatively ?
3. how to assess the uncertainties of the inverted models due to the interparameter tradeoffs ?

The first objective of this chapter is to evaluate the relative strengths and characteristics of interparameter contamination in isotropic-elastic FWI with multiparameter Hessian, which describes geometry of the objective function in terms of curvature or convexity (Fichtner and Trampert, 2011b; Fichtner and van Leeuwen, 2015). The diagonal blocks in the multiparameter Hessian characterize spatial correlations of the same physical parameter. Off-diagonal blocks measure correlations between different physical parameters (Fichtner and Trampert, 2011a; Operto et al., 2013). Rows in the multiparameter Hessian are averaging kernels (Backus and Gilbert, 1968) and columns are defined as multiparameter point spread functions (MPSFs) (Valenciano et al., 2006; Fichtner and Trampert, 2011b; Trampert et al., 2013; Tang and Lee, 2015; Zhu and Fomel, 2016). This chapter reveals that products of multiparameter Hessian off-diagonal blocks with the model perturbation vectors, which I will refer to as *interparameter contamination kernels*, account for the interparameter tradeoffs. For most large-scale inverse problems, explicitly constructing the Hessian matrix is considered to be computationally unaffordable. However, characteristics of the Hessian can be inferred via matrix probing techniques, which are useful when explicit representation of a matrix are too expensive to be constructed (Trampert et al., 2013). A low rank approximation of the Hessian can be efficiently computed by applying it to various types of vectors (Halko et al., 2011; Demanet et al., 2012; An, 2012; Zhu et al., 2016; Rawlinson and Spakman, 2016). This chapter also examines the adjoint-state and finite-difference approaches for multiparameter Hessian matrix-vector product calculation. The product of the Hessian with a point-localized model perturbation vector preserves one Hessian column (Spakman, 1991). The MPSFs measure the relative strengths and finite-frequency features of the local interparameter tradeoffs. Furthermore, it is also shown (see below in this chapter) that S-wave velocity perturbations tend generally to produce strong contaminations into density update and phase-reversed contaminations within the P-wave velocity update, which may make density highly under- or overestimated and cancel the update for P-wave velocity.

To assess the interparameter tradeoffs within the whole volume of interest, MPSFs should be computed for each type of model parameter at every spatial position, which also results in prohibitive computation expense (Fichtner and Trampert, 2011b; Chen and Xie, 2015). Assuming that the multiparameter Hessian matrix is diagonally dominant, I have adopted a stochastic probing strategy by applying multiparameter Hessian to random vectors. Expectation values of the correlations between the random vector with its Hessian-vector products approximate Hessian diagonals (Sacchi et al., 2007; MacCarthy et al., 2011; Trampert et al., 2013). Arranging different random probes, the diagonals of multiparameter Hessian off-diagonal blocks, which measure the coupling strengths of different physical parameters in the whole volume, can be estimated stochastically. Stochastic estimations of the Hessian diagonals can also be used as preconditioners for acceleration (Modrak and Tromp, 2016).

Reducing the uncertainties introduced by the interparameter tradeoff is becoming essential for multiparameter FWI. Newton-based optimization methods are promising because they incorporate the inverse multiparameter Hessian with its ability to suppress the unwanted parameter crosstalk artifacts (Innanen, 2014; Métivier et al., 2015; Wang et al., 2016; Yang et al., 2016). As stated in previous chapters, explicitly constructing and inverting multiparameter Hessian for large-scale inverse problems is, however, impracticably expensive as I have mentioned. Truncated-Newton (or Hessian-free) optimization methods represent affordable strategies for multiparameter FWI, in which the Newton equation is solved iteratively with matrix-free scheme of conjugate-gradient algorithm (Métivier et al., 2013; Boehm and Ulbrich, 2014; Métivier et al., 2015; Liu et al., 2015). However, iteratively solving the Newton equation is also expensive. Furthermore, by increasing savings through use of a small number of inner iterations, the effectiveness of removing interparameter mappings is reduced (Baumstein, 2014). Mode decomposition is a potential strategy for mitigating interparameter tradeoffs by isolating P and S wavefields but may also be limited in reducing the contaminations in density updates and multiparameter acoustic FWI (Wang and Cheng, 2017). Subspace optimization methods mitigate interparameter tradeoffs by scaling different physical parameters but do not prevent their occurrence (Kennett et al., 1988; Bernauer et al., 2014). In this chapter, based on a set of observations made on synthetic examples of interparameter tradeoff, a novel strategy is developed to reduce the interparameter tradeoff by approximating quantities I will refer to as *interparameter contamination kernels*. This strategy approximates the parameter contamination in model space by applying multiparameter Hessian off-diagonal blocks to estimated model vectors. The result is a model estimate which is approximately free of parameter crosstalk, and which has been created without iteratively solving large Newton systems, and which is in principle applicable to any tomographic or FWI misfit function. Numerical examples are given to illustrate that this new strategy is able to remove the contaminations from S-wave velocity partially and provide more reliable density estimations in isotropic-elastic FWI.

In addition to suppressing interparameter contaminations, parameter resolution quantification is key to a well-posed inverse scheme; it has been investigated by many researchers (Backus and Gilbert, 1968; Spakman, 1991; Fichtner and Trampert, 2011b; Rawlinson et al., 2014; Rawlinson and Spakman, 2016; Zhu et al., 2016). Within a Bayesian inference framework, uncertainties of the maximum a posterior model are evaluated based on posterior covariance operatorïijŇ which has a direct relationship with the (inverse) Hessian (Gouveia and Scales, 1998; Tarantola, 2005; Dettmer et al., 2007; Fichtner and Trampert, 2011b; Flath et al., 2011). In recent years, researchers have evaluated the local resolution of the inverted model with point spread functions by applying multiparameter Hessian to Gaussian-shape model perturbations (Fichtner and Trampert, 2011b; Rickers et al., 2013; Zhu et al., 2015; Bozdağ et al., 2016). However, the point spread functions actually represent conservative estimations of columns in resolution matrix by approximating inverse Hessian as an identity matrix (Oldenborger and Routh, 2009; Fichtner and van Leeuwen, 2015). The similarities and differences of the Hessian matrix and resolution matrix in resolution analysis are investigated. Approximate eigenvalue volumes are used to evaluate resolution of inverted models within the whole volume. Local spatial and interparameter tradeoffs of the inverted models are quantified with extended multiparameter point spread functions (EMPSFs) by applying the approximate inverse Hessian to the traditional MPSFs iteratively with preconditioned conjugate-gradient algorithm. The approximate inverse Hessian will de-blur the MPSFs further, balance the relative magnitudes by compensating geometrical spreading and mitigate interparameter contaminations, which represent more accurate local measurements of spatial and interparameter tradeoffs.

This chapter first reviews the basic principles of isotropic-elastic FWI. Benefits and limitations of the parameter resolution studies based upon scattering patterns are explored. *Interparameter contamination kernels* are then defined and how to evaluate the interparameter tradeoffs with multiparameter Hessian-vector products is explained. An explanation of the novel inversion strategy in reducing the interparameter tradeoffs with approximate contamination kernels is then given. Strategies for quantifying the resolution of the inverted models with approximate eigenvalue volumes and extended multiparameter point spread functions (EMPSFs) are explained. In the numerical modelling section, the strengths and characteristics of local interparameter tradeoffs are first examined with multiparameter point spread functions. Proposed matrix probing techniques are applied on a complex Marmousi model to assess the interparameter tradeoffs within the whole volume. The new inversion strategy is also applied to invert isotropic-elastic parameters with synthetic data and the low-frequency Hussar field seismic data set acquired by the CREWES Project and collaborators in 2011 (Margrave et al., 2012). The approximate eigenvalue volumes and EMPSFs are used to quantify resolution of the inverted models. The performances of various parameterizations for reconstructing subsurface isotropic-elastic properties are also examined with synthetic examples. Note: in this chapter, the expressions of gradients, Hessian-vector products, etc are given in time domain with integral formulas, whereas in previous chapters for convenience discrete formulas were used.

THEORY AND METHODS

Isotropic-elastic full-waveform inversion

The common l-2 norm misfit function in time domain can be expressed as:

$$\Phi(\mathbf{m}) = \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T \|\Delta\mathbf{d}(\mathbf{x}_s,\mathbf{x}_g,t;\mathbf{m})\|^2 dt, \tag{1}$$

where $\Delta\mathbf{d}(\mathbf{x}_s,\mathbf{x}_g,t;\mathbf{m}) = \mathbf{d}_{\text{syn}}(\mathbf{x}_s,\mathbf{x}_g,t) - \mathbf{d}_{\text{obs}}(\mathbf{x}_s,\mathbf{x}_g,t;\mathbf{m})$ is the data residual, $\mathbf{x}_s$ ($s = 1,...,S$) and $\mathbf{x}_g$ ($g = 1,...,R$) indicate source and receiver locations, S and R are the maximum source and receiver indexes, and T represents maximum recording time. In order to solve the inverse problem and find the model which minimizes the adopted cost function, the model is updated iteratively. The gradient of the misfit function can be obtained by correlating the Fréchet derivative wavefield with data residual:

$$\nabla_{\mathbf{m}}\Phi(\mathbf{m}) = \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T\int_{\Omega(\mathbf{x})}\nabla_{\mathbf{m}(\mathbf{x})}\mathbf{u}^{\star}(\mathbf{x}_s,\mathbf{x}_g,t;\mathbf{m})\,\Delta\mathbf{d}(\mathbf{x}_s,\mathbf{x}_g,t;\mathbf{m})\,d\mathbf{x}dt, \tag{2}$$

where $\nabla_{\mathbf{m}(\mathbf{x})}\mathbf{u}(\mathbf{x}_s,\mathbf{x}_g,t;\mathbf{m})$ indicates the Fréchet derivative wavefield, Ω indicates the whole volume, and the symbol $\star$ means complex conjugate transpose. Considering general anisotropic-elastic media, based on Born approximation the perturbed nth displacement field due to model perturbation $\Delta\mathbf{m}_\rho$ and $\Delta\mathbf{m}_{c_{ijkl}}$ is expressed as:

$$\begin{aligned}\Delta u_n(\mathbf{x}_s,\mathbf{x}_g,t;\Delta\mathbf{m}) = -\int_{\Omega(\mathbf{x})}\int_0^t &[\Delta m_\rho(\mathbf{x})\,G_{ni}(\mathbf{x},\mathbf{x}_r,t-t')\,\partial_t^2 u_i(\mathbf{x},\mathbf{x}_s,t') \\ &+ \Delta m_{c_{ijkl}}(\mathbf{x})\,\partial_j G_{ni}(\mathbf{x},\mathbf{x}_r,t-t')\,\partial_k u_l(\mathbf{x},\mathbf{x}_s,t')]dt'd\mathbf{x},\end{aligned} \tag{3}$$

where ρ and c_{ijkl} (i,j,k,l take on the values of x,y,z) denote density and elastic constant tensor with Einstein summation convention, G_{ni} is the Green's tensor, the nth displacement response due to impulse source at the ith direction. For isotropic-elastic media, the perturbation of elastic constants can be expressed in terms of the perturbations of bulk modulus $\Delta\mathbf{m}_\kappa$ and shear modulus $\Delta\mathbf{m}_\mu$: $\Delta m_{c_{ijkl}}(\mathbf{x}) = (\Delta m_\kappa(\mathbf{x}) - 2/3\Delta m_\mu(\mathbf{x}))\,\delta_{ij}\delta_{kl} + \Delta m_\mu(\mathbf{x})\,(\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il})$. Wavefield perturbation due to the perturbations of isotropic-elastic parameters are expressed as:

$$\begin{aligned}\Delta u_n(\mathbf{x}_s,\mathbf{x}_g,t;\Delta\mathbf{m}) = -\int_{\Omega(\mathbf{x})}\int_0^t &[\Delta m_\rho(\mathbf{x})\,G_{ni}(\mathbf{x},\mathbf{x}_r,t-t')\,\partial_t^2 u_i(\mathbf{x},\mathbf{x}_s,t') \\ &+ \left(\Delta m_\kappa(\mathbf{x}) - \frac{2}{3}\Delta m_\mu(\mathbf{x})\right)\delta_{ij}\delta_{kl}\partial_j G_{ni}(\mathbf{x},\mathbf{x}_r,t-t')\,\partial_k u_l(\mathbf{x},\mathbf{x}_s,t') \\ &+ \Delta m_\mu(\mathbf{x})\,(\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il})\,\partial_j G_{ni}(\mathbf{x},\mathbf{x}_r,t-t')\,\partial_k u_l(\mathbf{x},\mathbf{x}_s,t')]dt'd\mathbf{x}.\end{aligned} \tag{4}$$

Substituting equation (4) into equation (5) gives the Fréchet derivative of the misfit function:

$$\nabla_{\mathbf{m}}\Phi(\mathbf{m}) = \int_{\Omega(\mathbf{x})}[K_\kappa(\mathbf{x})\,a_\kappa(\mathbf{x}) + K_\mu(\mathbf{x})\,a_\mu(\mathbf{x}) + K_\rho(\mathbf{x})\,a_\rho(\mathbf{x})]d\mathbf{x}, \tag{5}$$

where K_κ, K_μ and K_ρ represent sensitivity kernels with respect to bulk modulus κ, shear modulus μ and density ρ, $a_\kappa = \Delta\mathbf{m}_\kappa/\mathbf{m}_\kappa$, $a_\mu = \Delta\mathbf{m}_\mu/\mathbf{m}_\mu$ and $a_\rho = \Delta\mathbf{m}_\rho/\mathbf{m}_\rho$ are relative model perturbations. Explicit expressions of the sensitivity kernels for these isotropic-elastic parameters can be written as (Tromp et al., 2005; Liu et al., 2006; Zhu et al., 2009; Luo et al., 2013; Yuan and Simons, 2014):

$$K_\kappa(\mathbf{x}) = -\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T \kappa(\mathbf{x})\,\partial_i\tilde{u}_i(\mathbf{x}_g,\mathbf{x},T-t)\,\partial_k u_k(\mathbf{x},\mathbf{x}_s,t)\,dt, \tag{6}$$

$$\begin{aligned} K_\mu(\mathbf{x}) = &-\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T \mu(\mathbf{x})\,[\partial_j\tilde{u}_i(\mathbf{x}_g,\mathbf{x},T-t)\,(\partial_i u_j(\mathbf{x},\mathbf{x}_s,t) + \partial_j u_i(\mathbf{x},\mathbf{x}_s,t)) \\ &-\frac{2}{3}\partial_i\tilde{u}_i(\mathbf{x}_g,\mathbf{x},T-t)\,\partial_k u_k(\mathbf{x},\mathbf{x}_s,t)]dt, \end{aligned} \tag{7}$$

$$K_\rho(\mathbf{x}) = -\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T \rho(\mathbf{x})\,\tilde{u}_i(\mathbf{x}_g,\mathbf{x},T-t)\,\partial_t^2 u_i(\mathbf{x},\mathbf{x}_s,t)\,dt, \tag{8}$$

where $\tilde{u}_i(\mathbf{x}_g,\mathbf{x},T-t)$ represents the ith component of the adjoint wavefield:

$$\tilde{u}_i(\mathbf{x}_g,\mathbf{x},T-t) = \int_0^{T-t} G_{in}(\mathbf{x}_g,\mathbf{x},T-t-t')\,\tilde{f}_n(\mathbf{x},t')\,dt', \tag{9}$$

where $\tilde{f}_n(\mathbf{x},t')$ is the adjoint source (Tromp et al., 2005; Bozdag et al., 2011):

$$\tilde{f}_n(\mathbf{x},t') = \sum_{\mathbf{x}_g}\Delta d_n(\mathbf{x}_g,T-t')\,\delta(\mathbf{x}-\mathbf{x}_g)\,. \tag{10}$$

With velocity-density parameterization, the corresponding sensitivity kernels for P-velocity α, S-wave velocity β and density ρ' are given by (Tromp et al., 2005; Köhn et al., 2012; Yuan et al., 2015):

$$K_\alpha = 2\left(1+\frac{4}{3}\frac{\mu}{\kappa}\right)K_\kappa, K_\beta = 2\left(K_\mu - \frac{4}{3}\frac{\mu}{\kappa}K_\kappa\right), K_{\rho'} = K_\rho + K_\kappa + K_\mu. \tag{11}$$

For impedance-density parameterization, sensitivity kernels of P-wave impedance IP=$\alpha\rho''$, S-wave impedance IS=$\beta\rho''$ and density ρ'' are given by:

$$\begin{aligned} K_{\mathrm{IP}} &= 2\left(1+\frac{4}{3}\frac{\mu}{\kappa}\right)K_\kappa = K_\alpha, \\ K_{\mathrm{IS}} &= 2\left(K_\mu - \frac{4}{3}\frac{\mu}{\kappa}K_\kappa\right) = K_\beta, \\ K_{\rho''} &= -K_\kappa - K_\mu + K_\rho = -K_\alpha - K_\beta + K_{\rho'}. \end{aligned} \tag{12}$$

Matrix multiplication of Newton equation system can be written with an integral formulation:

$$\nabla_{\mathbf{m}}\Phi(\mathbf{x}) = -\int_{\Omega(\mathbf{x}')} H(\mathbf{x},\mathbf{x}')\,\Delta m(\mathbf{x}')\,d\mathbf{x}', \tag{13}$$

where $H(\mathbf{x},\mathbf{x}')$ denotes one Hessian element described by positions $\mathbf{x}$ and $\mathbf{x}'$. In this chapter, to update the isotropic-elastic parameters simultaneously, a quasi-Newton l-BFGS optimization method is used. At each iteration, a line search approach is employed to obtain the step length for updating the model (Nocedal and Wright, 2006; Yuan et al., 2015).

Physical interpretation of multiparameter Hessian

In multiparameter FWI, the multiparameter Hessian has a block structure. For general anisotropic-elastic media, Hessian $\mathbf{H}$ can be expressed as:

$$\begin{aligned}\mathbf{H} = \int_{\Omega(\mathbf{x})}\int_{\Omega(\mathbf{x}')} & \left[\Delta m_\rho(\mathbf{x}) H_{\rho\rho}(\mathbf{x},\mathbf{x}') \Delta m_\rho(\mathbf{x}') + \Delta m_\rho(\mathbf{x}) H_{\rho\mathbf{c}}(\mathbf{x},\mathbf{x}') :: \Delta m_{\mathbf{c}}(\mathbf{x}')\right. \\ & \left. + \Delta m_{\mathbf{c}}(\mathbf{x}) :: H_{\mathbf{c}\rho}(\mathbf{x},\mathbf{x}') \Delta m_\rho(\mathbf{x}') + \Delta m_{\mathbf{c}}(\mathbf{x}) :: H_{\mathbf{c}\rho}(\mathbf{x},\mathbf{x}') :: \Delta m_{\mathbf{c}}(\mathbf{x}')\right] d\mathbf{x}d\mathbf{x}',\end{aligned} \tag{14}$$

where $\Delta\mathbf{m_c}$ indicates the perturbation of elastic constant tensor $\mathbf{c}$ and $\mathbf{::}$ means sequential contractions over the four nearest tensor indices (Luo, 2012). Because $\mathbf{H}$ is symmetric, then $\mathbf{H}_{\rho\mathbf{c}} = \mathbf{H}^{\dagger}_{\mathbf{c}\rho}$. Explicit expressions of diagonal blocks $\mathbf{H}_{\rho\rho}$ and $\mathbf{H}_{\mathbf{cc}}$ and off-diagonal block $\mathbf{H}_{\mathbf{c}\rho}$ are given by:

$$\begin{aligned}H_{\rho\rho}(\mathbf{x},\mathbf{x}') = & \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int\int \partial^2_{t'} u_i(\mathbf{x},\mathbf{x}_s,t') G_{ni}(\mathbf{x}_g,\mathbf{x},T-t') \\ & \times G_{n'i'}(\mathbf{x}_g,\mathbf{x}',t-t'') \partial^2_{t''} u_{i'}(\mathbf{x}',\mathbf{x}_s,t'') dt'dt'',\end{aligned} \tag{15}$$

$$\begin{aligned}H_{\mathbf{cc}}(\mathbf{x},\mathbf{x}') = & \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int\int \partial_k u_l(\mathbf{x},\mathbf{x}_s,t') \partial_j G_{ni}(\mathbf{x}_g,\mathbf{x},T-t') \\ & \times \partial_{j'} G_{n'i'}(\mathbf{x}_g,\mathbf{x}',t-t'') \partial_{k'} u_{l'}(\mathbf{x}',\mathbf{x}_s,t'') dt'dt'',\end{aligned} \tag{16}$$

$$\begin{aligned}H_{\mathbf{c}\rho}(\mathbf{x},\mathbf{x}') = & \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int\int \partial_k u_l(\mathbf{x},\mathbf{x}_s,t') \partial_j G_{ni}(\mathbf{x}_g,\mathbf{x},T-t') \\ & \times G_{n'i'}(\mathbf{x}_g,\mathbf{x}',t-t'') \partial^2_{t''} u_{i'}(\mathbf{x}',\mathbf{x}_s,t'') dt'dt''.\end{aligned} \tag{17}$$

For velocity-density parameterization in isotropic-elastic FWI, the Newton equation system for simultaneously updating P-wave velocity α, S-wave velocity β and density ρ' can be written as:

$$\begin{bmatrix}\mathbf{H}_{\alpha\alpha} & \mathbf{H}_{\alpha\beta} & \mathbf{H}_{\alpha\rho'} \\ \mathbf{H}_{\beta\alpha} & \mathbf{H}_{\beta\beta} & \mathbf{H}_{\beta\rho'} \\ \mathbf{H}_{\rho'\alpha} & \mathbf{H}_{\rho'\beta} & \mathbf{H}_{\rho'\rho'}\end{bmatrix}\begin{bmatrix}\Delta\mathbf{m}_\alpha \\ \Delta\mathbf{m}_\beta \\ \Delta\mathbf{m}_{\rho'}\end{bmatrix} = -\begin{bmatrix}\nabla_\alpha\Phi \\ \nabla_\beta\Phi \\ \nabla_{\rho'}\Phi\end{bmatrix}, \tag{18}$$

where $\nabla_\alpha\Phi$, $\nabla_\beta\Phi$, and $\nabla_{\rho'}\Phi$ are gradient vectors of α, β and ρ' respectively. Multiparameter Hessian elements can be classified into 4 types: (A) diagonal elements of the diagonal blocks account for geometrical spreading (i.e., $H_{\alpha\alpha}(\mathbf{x},\mathbf{x})$); (B) off-diagonal elements of the diagonal blocks measure the spatial correlations of model parameters with the same physical signature (i.e., $H_{\alpha\alpha}(\mathbf{x},\mathbf{x}')$ with $\mathbf{x} \neq \mathbf{x}'$); (C) diagonals of off-diagonal blocks indicate the strength of interparameter coupling at the same location (i.e., $H_{\alpha\beta}(\mathbf{x},\mathbf{x})$); (D) off-diagonal elements of off-diagonal blocks describe both spatial and interparameter trade-offs (i.e., $H_{\alpha\beta}(\mathbf{x},\mathbf{x}')$ with $\mathbf{x} \neq \mathbf{x}'$). One column of the multiparameter Hessian describes the blurring of an input delta function by the inverse operator, which is defined as multi-parameter point spread function (MPSF) (Fichtner and Trampert, 2011b; Trampert et al., 2013; Fichtner and van Leeuwen, 2015; Tang and Lee, 2015; Zhu and Fomel, 2016). For example, the column $\mathbf{H}_\beta(\mathbf{x},\mathbf{x}_N)$ indicates the correlation model parameter β at position $\mathbf{x}_N$ with model parameters α, β and ρ' at all positions in the whole volume.

Quantifying interparameter tradeoffs via multiparameter Hessian probing

First this section shows that the unwanted interparameter tradeoff artifacts can be described by *interparameter contamination kernels* defined as products of multiparameter Hessian off-diagonal blocks with model perturbation vectors. Thus, the interparameter tradeoffs in isotropic-elastic FWI can be quantified by probing the multiparameter Hessian with various test vectors.

Interparameter contamination kernels

Interparameter contamination kernels may be first introduced starting from standard sensitivity kernels. According to equation (11), the sensitivity kernel $K_\alpha(\mathbf{x})$ is written explicitly as:

$$K_\alpha(\mathbf{x}) = -\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T 2\rho'\alpha^2\partial_k u_k(\mathbf{x}_s,\mathbf{x},t)\int_0^{T-t}\partial_i G_{in}(\mathbf{x},\mathbf{x}_g,T-t-t')\,\tilde{f}_n(\mathbf{x},t')\,dt'dt, \tag{19}$$

where the adjoint source $\tilde{f}_n$ can be decomposed into three parts due to perturbations of $\Delta\mathbf{m}_\alpha$, $\Delta\mathbf{m}_\beta$ and $\Delta\mathbf{m}_{\rho'}$ respectively:

$$\tilde{f}_n(\mathbf{x},t';\Delta\mathbf{m}) = \tilde{f}_n(\mathbf{x},t';\Delta\mathbf{m}_\alpha) + \tilde{f}_n(\mathbf{x},t';\Delta\mathbf{m}_\beta) + \tilde{f}_n(\mathbf{x},t';\Delta\mathbf{m}_{\rho'}). \tag{20}$$

Ignoring multiple scattering components in the data residuals and following equations (4) and (10), the three adjoint sources in equation (20) can be expressed as:

$$\tilde{f}_{n'}(\mathbf{x},t';\Delta\mathbf{m}_\alpha) = \langle 2\rho'\alpha^2\partial_{i'}G_{i'n'}(\mathbf{x}')\,\Delta m_\alpha(\mathbf{x}')\,\partial_{k'}u_{k'}(\mathbf{x}')\,\delta(\mathbf{x}-\mathbf{x}_g)\rangle, \tag{21}$$

$$\begin{aligned}\tilde{f}_{n'}(\mathbf{x},t';\Delta\mathbf{m}_\beta) = &\langle 2\rho'\beta^2\left[\partial_{j'}G_{n'i'}(\mathbf{x}')\,\Delta m_\beta(\mathbf{x}')\left(\partial_{i'}u_{j'}(\mathbf{x}')+\partial_{j'}u_{i'}(\mathbf{x}')\right)\right.\\ &\left.-2\partial_{i'}G_{n'i'}(\mathbf{x}')\,\Delta m_\beta(\mathbf{x}')\,\partial_{k'}u_{k'}(\mathbf{x}')\right]\delta(\mathbf{x}-\mathbf{x}_g)\rangle,\end{aligned} \tag{22}$$

$$\begin{aligned}\tilde{f}_{n'}(\mathbf{x},t';\Delta\mathbf{m}_{\rho'}) = &\langle\rho'\left[\left(G_{n'i'}(\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\,\partial_{t'}^2u_{i'}(\mathbf{x}') + 2\alpha^2\partial_{i'}G_{n'i'}(\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\,\partial_{k'}u_{k'}(\mathbf{x}')\right)\right.\\ &+2\beta^2\left(\partial_{j'}G_{n'i'}(\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\left(\partial_{i'}u_{j'}(\mathbf{x}')+\partial_{j'}u_{i'}(\mathbf{x}')\right)\right.\\ &\left.\left.-2\partial_{i'}G_{n'i'}(\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\,\partial_{k'}u_{k'}(\mathbf{x}')\right)\right]\delta(\mathbf{x}-\mathbf{x}_g)\rangle,\end{aligned} \tag{23}$$

where the symbol $\langle\cdot\rangle$ indicates summation over sources, receivers, time and positions for sake of compactness. Inserting equations (21), (22) and (23) into equation (19) partitions the standard sensitivity kernel K_α into:

$$K_\alpha = K_{\alpha\leftrightarrow\alpha} + K_{\beta\to\alpha} + K_{\rho'\to\alpha}, \tag{24}$$

where the first term $K_{\alpha\leftrightarrow\alpha}$ represents the correct update kernel for α, and the second and third terms $K_{\beta\to\alpha}$ and $K_{\rho'\to\alpha}$ are defined as *interparameter contamination kernels* (Pan et al., 2017d,b,c), which represent the contaminations from β and ρ' to α:

$$K_{\alpha\leftrightarrow\alpha}(\mathbf{x}) = -\langle 2\rho'\alpha^2\partial_k u_k(\mathbf{x})\,\partial_i G_{in}(\mathbf{x})\left[2\rho'\alpha^2\partial_{i'}G_{i'n'}(\mathbf{x}')\,\Delta m_\alpha(\mathbf{x}')\,\partial_{k'}u_{k'}(\mathbf{x}')\right]\rangle, \tag{25}$$

$$K_{\beta\rightarrow\alpha}(\mathbf{x}) = -\left\langle 2\rho'\alpha^2\partial_k u_k(\mathbf{x})\,\partial_i G_{in}(\mathbf{x})\,2\rho'\beta^2\left[\partial_{j'}G_{n'i'}(\mathbf{x}')\,\Delta m_\beta(\mathbf{x}')\left(\partial_{i'}u_{j'}(\mathbf{x}')\right.\right.\right.$$
$$\left.\left.\left.+\partial_{j'}u_{i'}(\mathbf{x}')\right) - 2\partial_{i'}G_{n'i'}(\mathbf{x}')\,\Delta m_\beta(\mathbf{x}')\,\partial_{k'}u_{k'}(\mathbf{x}')\right]\right\rangle, \tag{26}$$

$$K_{\rho'\rightarrow\alpha}(\mathbf{x}) = -\left\langle 2\rho'\alpha^2\partial_k u_k(\mathbf{x})\,\partial_i G_{in}(\mathbf{x})\,\rho'\left[\left(G_{n'i'}(\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\,\partial_{t'}^2 u_{i'}(\mathbf{x}') + 2\alpha^2\partial_{i'}G_{n'i'}(\mathbf{x}')\right.\right.\right.$$
$$\times\Delta m_{\rho'}(\mathbf{x}')\,\partial_{k'}u_{k'}(\mathbf{x}')\Big) + 2\beta^2\left(\partial_{j'}G_{n'i'}(\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\left(\partial_{i'}u_{j'}(\mathbf{x}') + \partial_{j'}u_{i'}(\mathbf{x}')\right)\right.$$
$$\left.\left.\left.-2\partial_{i'}G_{n'i'}(\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\,\partial_{k'}u_{k'}(\mathbf{x}')\right)\right]\right\rangle. \tag{27}$$

Interparameter contamination kernels can also be explained and obtained with Newton equation (18). The gradient vector $\nabla_\alpha\Phi(\mathbf{x})$ in equation (18) can be written as an integral formulation:

$$\begin{aligned}\nabla_\alpha\Phi(\mathbf{x}) = &-\int_{\Omega(\mathbf{x}')} H_{\alpha\alpha}(\mathbf{x},\mathbf{x}')\,\Delta m_\alpha(\mathbf{x}')\,d\mathbf{x}' \\ &-\int_{\Omega(\mathbf{x}')} H_{\alpha\beta}(\mathbf{x},\mathbf{x}')\,\Delta m_\beta(\mathbf{x}')\,d\mathbf{x}' \\ &-\int_{\Omega(\mathbf{x}')} H_{\alpha\rho'}(\mathbf{x},\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\,d\mathbf{x}',\end{aligned} \tag{28}$$

where model perturbation vectors $\Delta\mathbf{m}_\beta$ and $\Delta\mathbf{m}_{\rho'}$ blurred by off-diagonal blocks $\mathbf{H}_{\alpha\beta}$ and $\mathbf{H}_{\alpha\rho'}$ in multiparameter Hessian are mapped into the update for parameter α. Equation (28) is equivalent to equation (24). Products of multiparameter Hessian block matrices with the model perturbation vectors are equivalent to the correct update kernel $K_{\alpha\leftrightarrow\alpha}$ and *interparameter contamination kernels* $K_{\beta\rightarrow\alpha}$ and $K_{\rho'\rightarrow\alpha}$ in equation (28). Similarly, gradient vectors $\nabla_\beta\Phi$ and $\nabla_{\rho'}\Phi$ can be written as:

$$\begin{aligned}\nabla_\beta\Phi(\mathbf{x}) &= a_\beta\left(K_{\alpha\rightarrow\beta}(\mathbf{x}) + K_{\beta\leftrightarrow\beta}(\mathbf{x}) + K_{\rho'\rightarrow\beta}(\mathbf{x})\right) \\ &= -\int_{\Omega(\mathbf{x}')} H_{\beta\alpha}(\mathbf{x},\mathbf{x}')\,\Delta m_\alpha(\mathbf{x}')\,d\mathbf{x}' \\ &\quad-\int_{\Omega(\mathbf{x}')} H_{\beta\beta}(\mathbf{x},\mathbf{x}')\,\Delta m_\beta(\mathbf{x}')\,d\mathbf{x}' \\ &\quad-\int_{\Omega(\mathbf{x}')} H_{\beta\rho'}(\mathbf{x},\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\,d\mathbf{x}',\end{aligned} \tag{29}$$

$$\begin{aligned}\nabla_{\rho'}\Phi(\mathbf{x}) &= a_{\rho'}\left(K_{\alpha\rightarrow\rho'}(\mathbf{x}) + K_{\beta\rightarrow\rho'}(\mathbf{x}) + K_{\rho'\leftrightarrow\rho'}(\mathbf{x})\right) \\ &= -\int_{\Omega(\mathbf{x}')} H_{\rho'\alpha}(\mathbf{x},\mathbf{x}')\,\Delta m_\alpha(\mathbf{x}')\,d\mathbf{x}' \\ &\quad-\int_{\Omega(\mathbf{x}')} H_{\rho'\beta}(\mathbf{x},\mathbf{x}')\,\Delta m_\beta(\mathbf{x}')\,d\mathbf{x}' \\ &\quad-\int_{\Omega(\mathbf{x}')} H_{\rho'\rho'}(\mathbf{x},\mathbf{x}')\,\Delta m_{\rho'}(\mathbf{x}')\,d\mathbf{x}',\end{aligned} \tag{30}$$

where $K_{\beta\leftrightarrow\beta}$ and $K_{\rho'\leftrightarrow\rho'}$ are correct update kernels for β and ρ', $K_{\alpha\rightarrow\beta}$ and $K_{\rho'\rightarrow\beta}$ described by off-diagonal blocks $\mathbf{H}_{\beta\alpha}$ and $\mathbf{H}_{\beta\rho'}$ indicate contaminations from α and ρ' to β, $K_{\alpha\rightarrow\rho'}$ and $K_{\beta\rightarrow\rho'}$ described by off-diagonal blocks $\mathbf{H}_{\rho'\alpha}$ and $\mathbf{H}_{\rho'\beta}$ are contaminations from α and β to ρ'. Explicit expressions of these correct update kernels and *interparameter contamination kernels* are given in Appendix .

According to equations (28), (29) and (30), gradient updates are linear combinations of correct model estimations and the contributions of the *interparameter contamination kernels*, which are determined by both of model perturbations and multiparameter Hessian off-diagonal blocks. In a generalized inversion framework, off-diagonal blocks of the multiparameter Hessian provide direct measurements of the parameter crosstalks, which are influenced by different wave modes, source-receiver illumination, parameterizations, etc. In large-scale inverse problems, it is always unaffordable to construct the whole Hessian matrix explicitly. One objective of this paper is to infer the characteristics of Hessian with matrix probing techniques by applying multiparameter Hessian to various types of vectors and quantify the interparameter tradeoffs in isotropic-elastic FWI. Products of multiparameter Hessian with an arbitrary vector can be calculated with adjoint-sate and finite-difference approaches, as explained in Appendix .

Multiparameter point spread functions

The multiparameter Hessian is first applied to model perturbation vector $\Delta\mathbf{m}$:

$$\Delta\mathbf{m} = \left[\Delta\mathbf{m}_\alpha = 0 \ \Delta\mathbf{m}_\beta = A_\beta \delta\left(\mathbf{x}-\mathbf{z}\right) \ \Delta\mathbf{m}_{\rho'} = 0\right]^\dagger, \tag{31}$$

where perturbations of P-wave velocity α and density ρ' are zeros and perturbation of S-wave velocity β is point located at position **z** with a strength of A_β. According to equations (28), (29) and (30), the correct update kernel for S-wave velocity $K_{\beta\leftrightarrow\beta}$ is given by:

$$K_{\beta\leftrightarrow\beta}\left(\mathbf{x},\mathbf{z}\right) = -a_\beta^{-1} A_\beta \int_{\Omega(\mathbf{x}')} H_{\beta\beta}\left(\mathbf{x},\mathbf{x}'\right)\delta\left(\mathbf{x}'-\mathbf{z}\right)d\mathbf{x}'. \tag{32}$$

According to the sifting property of delta function:

$$K_{\beta\leftrightarrow\beta}\left(\mathbf{x},\mathbf{z}\right) = -a_\beta^{-1} A_\beta \mathbf{H}_{\beta\beta}\left(\mathbf{x},\mathbf{z}\right). \tag{33}$$

Similarly, *interparameter contamination kernels* $K_{\beta\to\alpha}$ and $K_{\beta\to\rho'}$ are given by:

$$K_{\beta\to\alpha}\left(\mathbf{x},\mathbf{z}\right) = -a_\beta^{-1} A_\beta \int_{\Omega(\mathbf{x}')} H_{\alpha\beta}\left(\mathbf{x},\mathbf{x}'\right)\delta\left(\mathbf{x}'-\mathbf{z}\right)d\mathbf{x}' = -a_\beta^{-1} A_\beta \mathbf{H}_{\alpha\beta}\left(\mathbf{x},\mathbf{z}\right) \tag{34}$$

$$K_{\beta\to\rho'}\left(\mathbf{x},\mathbf{z}\right) = -a_\beta^{-1} A_\beta \int_{\Omega(\mathbf{x}')} H_{\rho'\beta}\left(\mathbf{x},\mathbf{x}'\right)\delta\left(\mathbf{x}'-\mathbf{z}\right)d\mathbf{x}' = -a_\beta^{-1} A_\beta \mathbf{H}_{\alpha\beta}\left(\mathbf{x},\mathbf{z}\right), \tag{35}$$

where $K_{\beta\to\alpha}\left(\mathbf{x},\mathbf{z}\right)$ and $K_{\beta\to\rho'}\left(\mathbf{x},\mathbf{z}\right)$ are local contaminations from β to α and ρ'. Multiparameter Hessian-vector product preserves the column of multiparameter Hessian $\mathbf{H}_\beta\left(\mathbf{x},\mathbf{z}\right) = \left[\mathbf{H}_{\alpha\beta}\left(\mathbf{x},\mathbf{z}\right) \ \ \mathbf{H}_{\beta\beta}\left(\mathbf{x},\mathbf{z}\right) \ \ \mathbf{H}_{\rho'\beta}\left(\mathbf{x},\mathbf{z}\right)\right]^\dagger$, which is referred to as a multiparameter point spread function (MPSF) following the common convention in exploration geophysics (Hu et al., 2001; Valenciano et al., 2006; Valenciano, 2008; Tang, 2009; Ren et al., 2011). Following equations (25), (26) and (27), the multiparameter point spread function $\mathbf{H}_\beta\left(\mathbf{x},\mathbf{z}\right)$ can be expressed explicitly as:

$$\begin{aligned}\mathbf{H}_{\beta\beta}\left(\mathbf{x},\mathbf{z}\right) =& \langle -2\rho'\beta^2\left(\mathbf{x}\right)\left[\partial_j G_{ni}\left(\mathbf{x}\right)\left(\partial_i u_j\left(\mathbf{x}\right)+\partial_j u_i\left(\mathbf{x}\right)\right)\right.\\ & \left. -2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right]\mathfrak{J}_\beta\left(\mathbf{z}\right)\rangle,\end{aligned} \tag{36}$$

$$\mathbf{H}_{\alpha\beta}\left(\mathbf{x},\mathbf{z}\right) = \langle -2\rho'\alpha^2\left(\mathbf{x}\right)\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\mathfrak{J}_\beta\left(\mathbf{z}\right)\rangle, \tag{37}$$

$$\begin{aligned}\mathbf{H}_{\rho'\beta}\left(\mathbf{x},\mathbf{z}\right) = \langle -\rho'\left(\mathbf{x}\right)\big[\big(G_{ni}\left(\mathbf{x}\right)\partial_t^2 u_i\left(\mathbf{x}\right) + 2\alpha^2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\big) \\ +2\beta^2\left(\partial_j G_{ni}\left(\mathbf{x}\right)\left(\partial_i u_j\left(\mathbf{x}\right)+\partial_j u_i\left(\mathbf{x}\right)\right) - 2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right)\big]\mathfrak{J}_\beta\left(\mathbf{z}\right)\rangle,\end{aligned} \tag{38}$$

where $\mathfrak{J}_\beta\left(\mathbf{z}\right)$ represents the product of Jacobian matrix due to parameter β with the point-localized model perturbation vector:

$$\begin{aligned}\mathfrak{J}_\beta\left(\mathbf{z}\right) = -\int_{\Omega(\mathbf{x}')} 2\rho'\beta^2\left(\mathbf{x}'\right)\left[\partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)\left(\partial_{i'}u_{j'}\left(\mathbf{x}'\right)+\partial_{j'}u_{i'}\left(\mathbf{x}'\right)\right)\right. \\ \left.-2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right]A_\beta\delta\left(\mathbf{x}'-\mathbf{z}\right)d\mathbf{x}'.\end{aligned} \tag{39}$$

Applying multiparameter Hessian to spike model perturbation $\Delta\mathbf{m}_\alpha = A_\alpha\delta\left(\mathbf{x}-\mathbf{z}\right)$ or $\Delta\mathbf{m}_{\rho'} = A_{\rho'}\delta\left(\mathbf{x}-\mathbf{z}\right)$ allows us to calculate the MPSFs $\mathbf{H}_{\beta\alpha}\left(\mathbf{x},\mathbf{z}\right)$, $\mathbf{H}_{\rho'\alpha}\left(\mathbf{x},\mathbf{z}\right)$, $\mathbf{H}_{\alpha\rho'}\left(\mathbf{x},\mathbf{z}\right)$, and $\mathbf{H}_{\beta\rho'}\left(\mathbf{x},\mathbf{z}\right)$, which describe the local contaminations from α to β and ρ' and the contaminations from ρ' to α and β. With these MPSFs, the relative strengths, phase characteristics and spreading widths of the local interparameter contaminations are evaluated by taking finite-frequency effects and source-receiver illumination into consideration. Because it is used within the context of the Born approximation, the amplitude of the spike model perturbation vector should be chosen to be smaller than 10% of the background model.

Evaluating interparameter tradeoffs within the whole volume

Multiparameter point spread functions (MPSFs) are limited in their ability to characterize the parameter resolution because they are spatially local. To evaluate the coupling effects of different physical parameters in the whole volume of interest, MPSFs would have to be computed for each type of model parameter at every spatial position, which gives rise to extensive computation requirements. An efficient stochastic probing approach is introduced to estimate the essential diagonals of subblock matrices in multiparameter Hessian. Diagonals of multiparameter Hessian off-diagonal blocks measure the coupling strengths of different physical parameters in the whole volume.

I first consider a function $v\left(\mathbf{x}\right)$, which satisfies $\mathbf{v} \sim \mathcal{N}\left(\mathbb{E}\left[\mathbf{v}\right], \mathbf{\Sigma_{vv}}\right)$ ($\mathcal{N}$ means Gaussian distribution). Expectation value $\mathbb{E}\left[\mathbf{v}\right]$ and variance-covariance matrix $\mathbf{\Sigma_{vv}}$ satisfy:

$$\mathbb{E}\left[v\left(\mathbf{x}\right)\right] = 0, \tag{40}$$

$$\begin{aligned}\Sigma_{\mathbf{vv}}\left(v\left(\mathbf{x}\right), v\left(\mathbf{x}'\right)\right) &= \mathbb{E}\left[\left(v\left(\mathbf{x}\right)-\mathbb{E}\left[v\left(\mathbf{x}\right)\right]\right)\left(v\left(\mathbf{x}'\right)-\mathbb{E}\left[v\left(\mathbf{x}'\right)\right]\right)^\dagger\right] \\ &= \mathbb{E}\left[v\left(\mathbf{x}\right)v\left(\mathbf{x}'\right)\right] - \mathbb{E}\left[v\left(\mathbf{x}\right)\right]\left(\mathbb{E}\left[v\left(\mathbf{x}'\right)\right]\right)^\dagger \\ &= \delta\left(\mathbf{x}-\mathbf{x}'\right).\end{aligned} \tag{41}$$

Correlating this random function with its Hessian-vector product $\mathfrak{H} = \mathbf{Hv}$ gives:

$$\begin{aligned}v\left(\mathbf{x}\right)\odot\mathfrak{H}\left(\mathbf{x}\right) &= \int_{\Omega(\mathbf{x}')} v\left(\mathbf{x}\right)H\left(\mathbf{x},\mathbf{x}'\right)v\left(\mathbf{x}'\right)d\mathbf{x}' \\ &= v\left(\mathbf{x}\right)H\left(\mathbf{x},\mathbf{x}\right)v\left(\mathbf{x}\right) + \int_{\Omega(\mathbf{x}')} v\left(\mathbf{x}\right)H_{\mathbf{x}\neq\mathbf{x}'}\left(\mathbf{x},\mathbf{x}'\right)v\left(\mathbf{x}'\right)d\mathbf{x}',\end{aligned} \tag{42}$$

where $\odot$ indicates element-wise multiplication, $H(\mathbf{x},\mathbf{x})$ and $H_{\mathbf{x}\neq\mathbf{x}'}(\mathbf{x},\mathbf{x}')$ represents Hessian diagonals and off-diagonals. Applying expectation operator $\mathbb{E}$ on both sides of equation (42) gives (Sacchi et al., 2007; Trampert et al., 2013):

$$\begin{aligned}
\mathbb{E}\left[v(\mathbf{x})\odot\mathfrak{H}(\mathbf{x})\right] &= \int_{\Omega(\mathbf{x}')} H(\mathbf{x},\mathbf{x}')\,\mathbb{E}\left[v(\mathbf{x})\,v(\mathbf{x}')\right]d\mathbf{x}' \\
&= \int_{\Omega(\mathbf{x}')} H(\mathbf{x},\mathbf{x}')\left(\Sigma_{\mathbf{vv}}\left(v(\mathbf{x}),v(\mathbf{x}')\right)+\mathbb{E}\left[v(\mathbf{x})\right]\left(\mathbb{E}\left[v(\mathbf{x}')\right]\right)^{\dagger}\right)d\mathbf{x}' \\
&= \int_{\Omega(\mathbf{x}')} H(\mathbf{x},\mathbf{x}')\,\delta(\mathbf{x}-\mathbf{x}')\,d\mathbf{x}' \\
&= H(\mathbf{x},\mathbf{x}),
\end{aligned} \tag{43}$$

where it can be seen that taking the expectation operation, the second term in equation (42), which represents off-diagonal elements, vanishes (Hutchinson, 1990; Trampert et al., 2013). The theoretical expectation operation can be approximated by averaging the cross-correlation results $\mathbf{v}\odot\mathfrak{H}$ with a finite number of independent zero-mean random vectors:

$$\mathbf{H}^{\text{diag}} \approx \sum_{nr=1}^{NR}\mathbf{v}_{nr}\odot\mathbf{H}\mathbf{v}_{nr}\oslash\sum_{nr}^{NR}\mathbf{v}_{nr}\odot\mathbf{v}_{nr} \tag{44}$$

where $\oslash$ represents element-wise division, nr is the index of random vector, NR indicates the maximum number of random vectors and $\mathbf{v}_{nr}\odot\mathbf{v}_{nr}$ is normalization term (MacCarthy et al., 2011). In a multiparameter inverse problem, the random vector $\mathbf{v}$ can be partitioned into N_p subvectors and multiparameter Hessian is divided into $N_p\times N_p$ subblock matrices, as illustrated in equation (18). Applying multiparameter Hessian to the random vector gives N_p sub-Hessian-vector products. Diagonals of the Hessian subblock matrices can be estimated by:

$$\mathbf{H}_{pq}^{\text{diag}} = \mathbb{E}\left[\mathbf{v}_p\odot\mathfrak{H}_p\right] = \mathbb{E}\left[\mathbf{v}_p\odot\sum_{q=1}^{N_p}\mathbf{H}_{pq}\mathbf{v}_q\right], \tag{45}$$

where p and q are indexes for subvectors representing different physical parameters, and $\mathfrak{H}_p$ is the sub-Hessian-vector product. Considering that zero-mean random vectors $\mathbf{v}_p$ and $\mathbf{v}_q$ for two different physical parameters are independent, equation (45) becomes:

$$\mathbf{H}_{pq}^{\text{diag}} = \sum_{q=1}^{N_p}\mathbf{H}_{pq}\mathbb{E}\left[\mathbf{v}_p\mathbf{v}_q\right] = \mathbf{H}_{pq}\mathbb{E}\left[\mathbf{v}_p\mathbf{v}_p\right]. \tag{46}$$

Proof of equation (46) is given in Appendix . With a series of random vectors, diagonals of the Hessian subblock matrices can be obtained approximately by:

$$\mathbf{H}_{pq}^{\text{diag}} \approx \sum_{nr=1}^{NR}\mathbf{v}_{p,nr}\odot\mathbf{H}_{pq}\mathbf{v}_{p,nr}\oslash\sum_{nr=1}^{NR}\mathbf{v}_{p,nr}\odot\mathbf{v}_{p,nr}. \tag{47}$$

In isotropic-elastic FWI, the random vector is given by $\mathbf{v}=\left[\mathbf{v}_\alpha\;\mathbf{v}_\beta\;\mathbf{v}_{\rho'}\right]^{\dagger}$, where $\mathbf{v}_\alpha$, $\mathbf{v}_\beta$, and $\mathbf{v}_{\rho'}$ are independent zero-mean subvectors. Applying multiparameter Hessian to this random vector gives three sub-Hessian-vector products $\mathfrak{H}_\alpha$, $\mathfrak{H}_\beta$ and $\mathfrak{H}_{\rho'}$:

$$\mathfrak{H}_\alpha = \mathbf{H}_{\alpha\alpha}\mathbf{v}_\alpha+\mathbf{H}_{\alpha\beta}\mathbf{v}_\beta+\mathbf{H}_{\alpha\rho'}\mathbf{v}_{\rho'}, \tag{48}$$

$$\mathfrak{H}_{\beta} = \mathbf{H}_{\beta\alpha}\mathbf{v}_{\alpha} + \mathbf{H}_{\beta\beta}\mathbf{v}_{\beta} + \mathbf{H}_{\beta\rho'}\mathbf{v}_{\rho'}, \tag{49}$$

$$\mathfrak{H}_{\rho'} = \mathbf{H}_{\rho'\alpha}\mathbf{v}_{\alpha} + \mathbf{H}_{\rho'\beta}\mathbf{v}_{\beta} + \mathbf{H}_{\rho'\rho'}\mathbf{v}_{\rho'}. \tag{50}$$

With a series of independent zero-mean random vectors, diagonals of the Hessian sub-blocks $\mathbf{H}_{\alpha\alpha}$ and $\mathbf{H}_{\beta\alpha}$ can be estimated approximately by:

$$\begin{aligned}\mathbf{H}^{\text{diag}}_{\alpha\alpha} &= \mathbb{E}\left[\mathbf{v}_{\alpha} \odot \mathfrak{H}_{\alpha}\right] \\ &= \mathbb{E}\left[\mathbf{v}_{\alpha} \odot \mathbf{H}_{\alpha\alpha}\mathbf{v}_{\alpha}\right] + \mathbb{E}\left[\mathbf{v}_{\alpha} \odot \mathbf{H}_{\alpha\beta}\mathbf{v}_{\beta}\right] + \mathbb{E}\left[\mathbf{v}_{\alpha} \odot \mathbf{H}_{\alpha\rho'}\mathbf{v}_{\rho'}\right] \\ &\approx \sum_{nr=1}^{NR} \mathbf{v}_{\alpha,nr} \odot \mathbf{H}_{\alpha\alpha}\mathbf{v}_{\alpha,nr} \oslash \sum_{nr=1}^{NR} \mathbf{v}_{\alpha,nr} \odot \mathbf{v}_{\alpha,nr},\end{aligned} \tag{51}$$

$$\begin{aligned}\mathbf{H}^{\text{diag}}_{\beta\alpha} &= \mathbb{E}\left[\mathbf{v}_{\alpha} \odot \mathfrak{H}_{\beta}\right] \\ &= \mathbb{E}\left[\mathbf{v}_{\alpha} \odot \mathbf{H}_{\beta\alpha}\mathbf{v}_{\alpha}\right] + \mathbb{E}\left[\mathbf{v}_{\alpha} \odot \mathbf{H}_{\beta\beta}\mathbf{v}_{\beta}\right] + \mathbb{E}\left[\mathbf{v}_{\alpha} \odot \mathbf{H}_{\beta\rho'}\mathbf{v}_{\rho'}\right] \\ &\approx \sum_{nr=1}^{NR} \mathbf{v}_{\alpha,nr} \odot \mathbf{H}_{\beta\alpha}\mathbf{v}_{\alpha,nr} \oslash \sum_{nr=1}^{NR} \mathbf{v}_{\alpha,nr} \odot \mathbf{v}_{\alpha,nr}.\end{aligned} \tag{52}$$

A similar approach can be used to estimate the diagonals of $\mathbf{H}_{\beta\beta}$, $\mathbf{H}_{\rho'\rho'}$, $\mathbf{H}_{\alpha\rho'}$ and $\mathbf{H}_{\beta\rho'}$. The choice of maximum random vectors NR depends on the desired accuracy of the estimated diagonals, which can be evaluated by statistically examining repeated estimates with independent random vectors (MacCarthy et al., 2011). Generally, more random probes give better estimations. If the sublocks of multiparameter Hessian are diagonally dominant, much less random probes are needed (Trampert et al., 2013). Sacchi et al. (2007) estimated the diagonal Hessian preconditioner with 5 random realizations using a phase shift approach. In this chapter, we show that diagonals of multiparameter Hessian can be estimated stochastically with 1 or 2 random Hessian-vector applications using spectral-element method. Stochastic estimations of Hessian diagonals can also be used as effective preconditioners in the inversion process.

Reducing interparameter tradeoffs with approximate contamination kernels

Non-uniqueness due to interparameter tradeoffs will increase nonlinearity and uncertainties within multiparameter inverse problems significantly. Different strategies including Newton-based optimization methods (Métivier et al., 2015; Liu et al., 2015), subspace optimization methods (Kennett et al., 1988; Baumstein, 2014; Bernauer et al., 2014), and wave mode decomposition strategies (Wang and Cheng, 2017), have been proposed to reduce the influences of interparameter tradeoffs in multiparameter FWI. However, most of these strategies have some limitations, as discussed in introduction section. In the numerical modelling section, with the proposed probing strategies, I find that S-wave velocity dominates the inversion process and produces relatively strong contaminations into density and P-wave velocity updates but suffers very weak contaminations from other parameters. Based on these observations, a novel inversion strategy has been developed to reduce the contaminations from S-wave velocity to other parameters especially density by approximating the contamination kernels.

From equation (30), it can be seen that the standard sensitivity kernel $K_{\rho'}$ is just linear summation of the correct update kernel $K_{\rho'\leftrightarrow\rho'}$ with two contamination kernels $K_{\alpha\rightarrow\rho'}$ and

$K_{\beta\rightarrow\rho'}$. The interparameter mappings from α and β to ρ' can be removed completely by simply summing the Hessian-vector products $\mathfrak{H}_{\rho'\alpha} = \mathbf{H}_{\rho'\alpha}\Delta\mathbf{m}_\alpha$ and $\mathfrak{H}_{\rho'\beta} = \mathbf{H}_{\rho'\beta}\Delta\mathbf{m}_\beta$ with standard sensitivity kernel $K_{\rho'}$. However, true model perturbation vectors $\Delta\mathbf{m}_\alpha$ and $\Delta\mathbf{m}_\beta$ are unknown variables. Because S-wave velocity suffers little contaminations from other parameters, the model parameters can be updated simultaneously for a finite number of k' iterations and then the inverted P-wave velocity and density models are dropped. The estimated S-wave velocity $\mathbf{m}_\beta^{k'}$ is kept. The inversion is then started from initial models by simultaneously updating three model parameters. At the $\tilde{k}$th iteration, the approximate contamination kernels $\tilde{K}^{\tilde{k}}_{\beta\rightarrow\alpha}$ and $\tilde{K}^{\tilde{k}}_{\beta\rightarrow\rho'}$ are constructed:

$$\tilde{K}^{\tilde{k}}_{\beta\rightarrow\alpha}(\mathbf{x}) = -\int_{\Omega(\mathbf{x}')} H^{\tilde{k}}_{\alpha\beta}(\mathbf{x},\mathbf{x}')\,\Delta\tilde{m}^{\tilde{k}}_\beta(\mathbf{x}')\,d\mathbf{x}', \tag{53}$$

$$\tilde{K}^{\tilde{k}}_{\beta\rightarrow\rho'}(\mathbf{x}) = -\int_{\Omega(\mathbf{x}')} H^{\tilde{k}}_{\rho'\beta}(\mathbf{x},\mathbf{x}')\,\Delta\tilde{m}^{\tilde{k}}_\beta(\mathbf{x}')\,d\mathbf{x}', \tag{54}$$

where $\Delta\tilde{\mathbf{m}}^{\tilde{k}}_\beta = \mathbf{m}^{k'}_\beta - \mathbf{m}^{\tilde{k}}_\beta$ is the approximate model perturbation vector. Subtracting the approximate contamination kernels from the standard sensitivity kernels $K^{\tilde{k}}_\alpha$ and $K^{\tilde{k}}_{\rho'}$ will remove the contaminations partially and give the new update kernels for α, β and ρ':

$$\tilde{K}^{\tilde{k}}_\alpha(\mathbf{x}) = K^{\tilde{k}}_\alpha(\mathbf{x}) - \tilde{K}^{\tilde{k}}_{\beta\rightarrow\alpha}(\mathbf{x}),\, \tilde{K}^{\tilde{k}}_\beta(\mathbf{x}) = K^{\tilde{k}}_\beta(\mathbf{x}),\, \tilde{K}^{\tilde{k}}_{\rho'}(\mathbf{x}) = K^{\tilde{k}}_{\rho'}(\mathbf{x}) - \tilde{K}^{\tilde{k}}_{\beta\rightarrow\rho'}(\mathbf{x}), \tag{55}$$

in which the S-wave velocity kernel $\tilde{K}^{\tilde{k}}_\beta$ is kept unchanged. A better approximation of the model perturbation vector $\Delta\tilde{\mathbf{m}}_\beta$ removes the contaminations more completely but at the cost of more computation requirements. Table 4.1 illustrates the basic work-flow for this inversion strategy. In traditional simultaneous inversion strategy, the computational cost of $\tilde{k}_{max}$ iterations is equivalent to number of $2 \times N_s \times \tilde{k}_{max}$ forward and adjoint simulations. This new inversion strategy will be more expensive for obtaining $\mathbf{m}^{k'}_\beta$ and constructing approximate contamination kernels. For $\tilde{k}_{max}$ iterations, the number of forward and adjoint simulations is equivalent to $\left(2 \times N_s \times k' \times N_{k'} + 8 \times N_s \times \tilde{k}_{max}\right)$, where $N_{k'}$ is number of loops for obtaining $\mathbf{m}^{k'}_\beta$.

Resolution analysis

Resolution analysis is long lasting issue for geophysical inverse problems and have been studied by many researchers (Backus and Gilbert, 1968; Spakman, 1991; Rawlinson et al., 2014; Rawlinson and Spakman, 2016). Quantifying resolution and uncertainties of the inverted models due to interparameter tradeoffs is a key aspect of for multiparameter FWI. Assuming that an optimal model $\mathbf{m}$ has been obtained with least-squares optimization framework, applying model perturbation $\Delta\mathbf{m}$ gives perturbed model $\mathbf{m}' = \mathbf{m} + \Delta\mathbf{m}$, which is close to model $\mathbf{m}$. The reconstructed model $\tilde{\mathbf{m}}$ can be obtained by $\tilde{\mathbf{m}} = \mathbf{m} + \Delta\mathbf{m} + \Delta\tilde{\mathbf{m}}$, where $\Delta\tilde{\mathbf{m}}$ represents the estimated model perturbation vector:

$$\Delta\tilde{\mathbf{m}} = -\mathbf{H}^{-g}\nabla_{\mathbf{m}}\Phi = \mathbf{H}^{-g}\mathbf{H}\Delta\mathbf{m} = \mathbf{R}\Delta\mathbf{m}, \tag{56}$$

where $\mathbf{H}^{-g}$ is the generalized inverse of $\mathbf{H}$ and $\mathbf{R} = \mathbf{H}^{-g}\mathbf{H}$ is the resolution matrix, which describes how the estimated model perturbation $\Delta\tilde{\mathbf{m}}$ relates to the true model perturbation

Notations: $\tilde{k}_{max}$ is the maximum iteration; ϕ is the normalized misfit; ϕ_{min} is the minimum normalized misfit; $\mathbf{N}_f$ is the frequency band.

Input: $\leftarrow \mathbf{m}^0 = [\mathbf{m}^0_\alpha \; \mathbf{m}^0_\beta \; \mathbf{m}^0_{\rho'}]^\dagger$, ϕ_{min}, $\tilde{k}_{max}$, $\tilde{\tilde{k}}_{max}$, k', $\mathbf{N}_f$, $\mathbf{d}_{\text{obs}}$

Output: $\rightarrow \mathbf{m}^{\text{est}}$, ϕ

Initialization: $\tilde{k} = 0$

For $\tilde{k} < \tilde{k}_{max}$ or $\phi_{\tilde{k}} > \phi_{min}$

1. $\leftarrow \mathbf{m}^{\tilde{k}}$, k', $\mathbf{N}_f$, $\mathbf{d}_{\text{obs}}$ \\ Iteratively estimate β by k' iterations $\rightarrow \mathbf{m}^{k'}_\beta$;
2. **For** $\tilde{\tilde{k}} < \tilde{\tilde{k}}_{max}$

   2.1. $\leftarrow \mathbf{m}^{\tilde{k}}$, $\mathbf{N}_f$, $\mathbf{d}_{\text{obs}}$ \\ Calculate sensitivity kernels $\rightarrow K^{\tilde{\tilde{k}}}_\alpha$, $K^{\tilde{\tilde{k}}}_\beta$ and $K^{\tilde{\tilde{k}}}_{\rho'}$;

   2.2. $\leftarrow \Delta\tilde{\mathbf{m}}^{\tilde{\tilde{k}}}_\beta = \mathbf{m}^{k'}_\beta - \mathbf{m}^{\tilde{\tilde{k}}}_\beta$ \\ Calculate approximate contamination kernels:
   $\rightarrow \tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\alpha} = -\mathbf{H}^{\tilde{\tilde{k}}}_{\alpha\beta}\Delta\tilde{\mathbf{m}}^{\tilde{\tilde{k}}}_\beta$, $\tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\rho'} = -\mathbf{H}^{\tilde{\tilde{k}}}_{\rho'\beta}\Delta\tilde{\mathbf{m}}^{\tilde{\tilde{k}}}_\beta$

   2.3. $\leftarrow \tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\alpha}$ and $\tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\rho'}$ \\ Calculate new update kernels:
   $\rightarrow \tilde{K}^{\tilde{\tilde{k}}}_\alpha = K^{\tilde{\tilde{k}}}_\alpha - \tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\alpha}$, $\tilde{K}^{\tilde{\tilde{k}}}_\beta = K^{\tilde{\tilde{k}}}_\beta$, $\tilde{K}^{\tilde{\tilde{k}}}_{\rho'} = K^{\tilde{\tilde{k}}}_{\rho'} - \tilde{K}^{\tilde{\tilde{k}}}_{\beta\rightarrow\rho'}$

   2.4. Apply stochastic estimations of diagonal Hessian preconditioners:

   2.5. Get step length $\mu_{\tilde{\tilde{k}}}$ with a line search method;

   2.6. Update the model vector: $\mathbf{m}^{\tilde{\tilde{k}}+1} = \mathbf{m}^{\tilde{\tilde{k}}} + \mu_{\tilde{\tilde{k}}}\Delta\mathbf{m}^{\tilde{\tilde{k}}}$;

   2.7. Calculate misfit $\phi_{\tilde{\tilde{k}}}$ and $\tilde{\tilde{k}} = \tilde{\tilde{k}} + 1$;

   End

3. Update parameters: $\tilde{k} = \tilde{k} + \tilde{\tilde{k}}_{max}$, $\phi_{\tilde{k}} = \phi_{\tilde{\tilde{k}}}$, $\mathbf{m}^{\text{est}} = \mathbf{m}^{\tilde{k}} = \mathbf{m}^{\tilde{\tilde{k}}_{max}}$;

End

Table 1. Work-flow of the new inversion strategy for isotropic-elastic FWI with approximate contamination kernels.

$\Delta\mathbf{m}$. Ideally $\mathbf{R}$ should be an identity matrix $\mathbf{I}$ meaning that the model is perfectly recovered (Backus and Gilbert, 1968). However, if the resolution matrix deviates significantly from the identity matrix, the inverted model suffers from tradeoffs (Luo, 2012). A column of $\mathbf{R}$ measures the local resolution and uncertainties of the inverted model (Oldenborger and Routh, 2009; Fichtner and Trampert, 2011b). However, explicitly constructing and inverting $\mathbf{H}$ are computationally unaffordable for large-scale inverse problems. In recent years, researchers evaluated the local resolution of the inverted model with point spread functions by approximating $\mathbf{H}^{-\mathrm{g}}$ with an identity matrix $\mathbf{I}$ (Fichtner and Trampert, 2011b; Rickers et al., 2013; Zhu et al., 2015; Bozdağ et al., 2016). Thus, the column of multiparameter Hessian (i.e., $\mathbf{H}(\mathbf{x},\mathbf{z})$ equation (33)) only represents an conservative estimation of the column in resolution matrix (i.e., $\mathbf{R}(\mathbf{x},\mathbf{z})$) (Fichtner and van Leeuwen, 2015). In this section, the similarities and differences between $\mathbf{H}(\mathbf{x},\mathbf{z})$ and $\mathbf{R}(\mathbf{x},\mathbf{z})$ in resolution analysis are investigated and the potential benefits by applying approximate inverse Hessian operators to PSFs are explored.

The symmetric and positive semi-definite Hessian matrix $\mathbf{H}$ can be decomposed as:

$$\mathbf{H} = \Xi\Pi\Xi^{-1}, \tag{57}$$

where $\Xi = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, ..., \mathbf{a}_M]$ is an orthogonal matrix consisting of M column eigenvectors $\mathbf{a}_{\dot{k}}, \dot{k} = 1, 2, ..., M$ of $\mathbf{H}$ and Π is a diagonal matrix with corresponding eigenvalues $\lambda_{\dot{k}}$. The generalized inverse of $\mathbf{H}$ is given by:

$$\mathbf{H}^{-g} = \left[\Xi\left(\Pi + \lambda_0\mathbf{I}\right)\Xi^{-1}\right]^{-1} = \Xi\left(\Pi + \lambda_0\mathbf{I}\right)^{-1}\Xi^{-1}, \tag{58}$$

where $\lambda_0\mathbf{I}$ is the damping term added to the eigenvalues. The resolution matrix $\mathbf{R}$ can be obtained by:

$$\mathbf{R} = \Xi\tilde{\Pi}^{-1}\Pi\Xi^{-1} = \Xi\left(\Pi + \lambda_0\mathbf{I}\right)^{-1}\Pi\Xi^{-1}, \tag{59}$$

where $\tilde{\Pi} = \left(\Pi + \lambda_0\mathbf{I}\right)^{-1}\Pi$ is diagonal matrix with eigenvalues of $\tilde{\lambda} = \left(1 + \lambda_0/\lambda_{\dot{k}}\right)^{-1}$. The Hessian matrix and resolution matrix have the same eigenvectors but different eigenvalues. Because the orthogonal eigenvectors of $\mathbf{H}$ span the model space, the model perturbation vector $\Delta\mathbf{m}$ can also be written as a sum of M eigenvectors $\mathbf{a}_{\dot{k}}$:

$$\Delta\mathbf{m} = \sum_{\dot{k}=1}^{M} h_{\dot{k}}\mathbf{a}_{\dot{k}} = h_1\mathbf{a}_1 + h_2\mathbf{a}_2 + ... + h_M\mathbf{a}_M, \tag{60}$$

where $h_{\dot{k}}$ are the model expansion coefficients. Combining equation (60) and equation (57), Hessian-vector product $\mathbf{H}\Delta\mathbf{m}$ can be expressed in terms of eigenvalues and eigenvectors of $\mathbf{H}$:

$$\mathbf{H}\Delta\mathbf{m} = \sum_{\dot{k}=1}^{M} \lambda_{\dot{k}} h_{\dot{k}}\mathbf{a}_{\dot{k}} = \lambda_1 h_1\mathbf{a}_1 + \lambda_2 h_2\mathbf{a}_2 + ... + \lambda_M h_M\mathbf{a}_M. \tag{61}$$

Substituting equations (60) and (59) into equation (56) gives:

$$\begin{aligned}\tilde{\mathbf{m}} &= \sum_{\dot{k}=1}^{M} \left(1 + \lambda_0/\lambda_{\dot{k}}\right)^{-1} h_{\dot{k}}\mathbf{a}_{\dot{k}}, \\ &= \left(1 + \lambda_0/\lambda_1\right)^{-1} h_1\mathbf{a}_1 + \left(1 + \lambda_0/\lambda_2\right)^{-1} h_2\mathbf{a}_2 + ... + \left(1 + \lambda_0/\lambda_M\right)^{-1} h_M\mathbf{a}_M.\end{aligned} \tag{62}$$

Assuming that the eigenvalues of $\mathbf{H}$ are constant $\lambda_{\dot{k}} \approx \lambda$, equations (61) and (62) become:

$$\begin{aligned}\mathbf{H}\Delta\mathbf{m} &\approx \sum_{\dot{k}=1}^{M} \lambda h_{\dot{k}} \mathbf{a}_{\dot{k}} = \lambda\Delta\mathbf{m}, \\ \tilde{\mathbf{m}} &\approx \sum_{\dot{k}=1}^{M} \left(1+\lambda_0/\lambda\right)^{-1} h_{\dot{k}} \mathbf{a}_{\dot{k}} = \left(1+\lambda_0/\lambda\right)^{-1} \Delta\mathbf{m}.\end{aligned} \tag{63}$$

Magnitudes of the PSFs directly measure the magnitudes of eigenvalues. Larger eigenvalues mean well constrained eigenvectors. Smaller eigenvalues mean poorly constrained eigenvectors. PSFs determine resolution of the inverted models with eigenvalues and mimic the shape of the true model perturbation but have distinct magnitudes. Because $\lambda_{\dot{k}} \gg \lambda_0$, then $\left(1+\lambda_0/\lambda_{\dot{k}}\right)^{-1} \approx 1$, the closer of eigenvalues of the resolution matrix approach 1, the better of the resolution is. If the Hessian is diagonally dominant, eigenvalues of the resolution matrix within the whole volume can be approximated by diagonals of the resolution matrix (Luo, 2012; Zhu et al., 2015), referred to as approximate eigenvalue volume:

$$\mathbf{Eig} = \left[\mathbf{H}^{\text{diag}} + \tilde{\epsilon} \times \max\left(\mathbf{H}^{\text{diag}}\right)\right]^{-1} \mathbf{H}^{\text{diag}}, \tag{64}$$

where $\max\left(\mathbf{H}^{\text{diag}}\right)$ represents the maximum value of diagonal Hessian $\mathbf{H}^{\text{diag}}$ and $\tilde{\epsilon}$ is a small constant value.

Here, this chapter proposes to quantify the local spatial and interparameter tradeoffs of the inverted models with extended multiparameter point spread functions (EMPSFs) by applying approximate inverse Hessian to MPSFs with conjugate-gradient algorithm preconditioned by stochastic estimations of diagonal Hessian. Considering a point-localized model perturbation vector $\Delta\mathbf{m} = [\Delta\mathbf{m}_\alpha = 0 \ \Delta\mathbf{m}_\beta = A_\beta\delta\left(\mathbf{x}-\mathbf{z}\right) \ \Delta\mathbf{m}_{\rho'} = 0]^\dagger$, equation (56) can be written as:

$$\Delta\tilde{m}_\beta\left(\mathbf{z}\right) = \int_{\Omega(\mathbf{x})}\int_{\Omega(\mathbf{x}')} A_\beta \tilde{R}_\beta\left(\mathbf{x},\mathbf{x}'\right)\delta\left(\mathbf{x}'-\mathbf{z}\right) d\mathbf{x}' d\mathbf{x} = \int_{\Omega(\mathbf{x})} A_\beta \tilde{R}_\beta\left(\mathbf{x},\mathbf{z}\right) d\mathbf{x}, \tag{65}$$

where $\tilde{\mathbf{R}}_\beta\left(\mathbf{x},\mathbf{z}\right) = \mathbb{H}^{-1}\left(\mathbf{H}_\beta\left(\mathbf{x},\mathbf{z}\right)\right)$ indicates the extended MPSF (EMPSF) and $\mathbb{H}^{-1}$ represents the approximate inverse Hessian by preconditioned conjugate-gradient algorithm. Applying the inverse Hessian approximately will re-scale the magnitudes and de-blur the MPSFs. Furthermore, approximate inverse multiparameter Hessian will also suppress the interparameter contaminations to a certain extent (Innanen, 2014; Métivier et al., 2015; Pan et al., 2016; Wang and Cheng, 2017). Thus, the EMPSFs will provide more accurate measurements of the local spatial and interparameter tradeoffs. To evaluate the interparameter tradeoffs of the inverted models obtained by new inversion strategy, the EMPSFs with a variant of preconditioned conjugate-gradient approach are constructed following the work-flow illustrated in Table 4.1.

NUMERICAL EXAMPLES

In the numerical modelling section, the proposed strategies are applied to quantify and reduce the interparameter tradeoffs in isotropic-elastic FWI. Spectral-element methods are employed for forward and adjoint simulations with the open-source software package *SPECFEM2D* (Komatitsch and Tromp, 2005). Influences of surface waves are currently not considered in the numerical examples presented in this chapter.

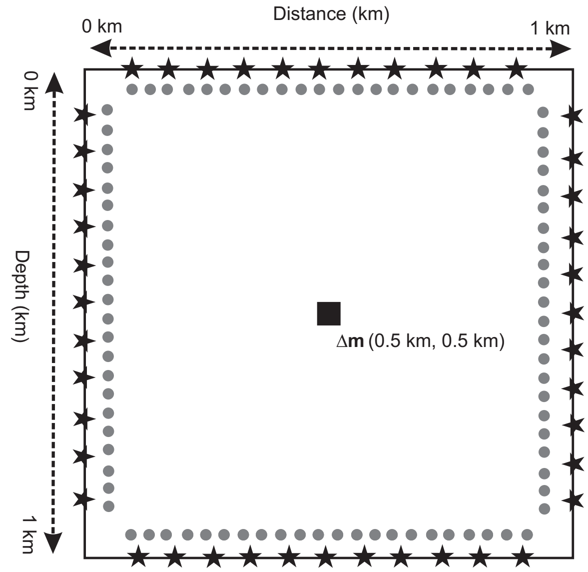

FIG. 1. Acquisition geometry for spike probing test. Black stars and gray circles represent sources and receivers positions. The 2D model is discretized into 50 and 50 uniform mesh nodes in horizontal and vertical directions with 1 km in width and 1 km in depth. The black square located at the center of the model indicates the spike model perturbation at position $\mathbf{z} = (0.5\ \text{km}, 0.5\ \text{km})$.

Spike probing test with MPSFs

The relative strengths and characteristics of the interparameter contaminations are first investigated with multiparameter point spread functions (MPSFs) in isotropic-elastic FWI using x-z component data. Inversion experiments with Gaussian-anomaly examples are given to verify the predictions and examine the effectiveness of this new inversion strategy in reducing the interparameter contaminations.

Figure 1 shows the 2D isotropic-elastic model with one spike model perturbation embedded in a homogeneous background. P-wave velocity, S-wave velocity and density of the background model are 2.0 km/s, 1.4 km/s and 1.2 g/cm^3. A P-SV mode source with Ricker wavelet (dominant frequency f_{dom}=8Hz) is used for modeling. A total of 60 sources and 200 receivers are arranged along all boundaries of the model with a regular source spacing of 62.5 m and a regular receiver spacing of 20 m. I first apply a positive spike model perturbation of P-wave velocity at position $\mathbf{z}$ (the model center): $\Delta\mathbf{m}_\alpha(\mathbf{z}) = 0.1$ km/s. Multiparameter point spread functions (MPSFs) $\mathbf{H}_{\alpha\alpha}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\alpha}(\mathbf{x},\mathbf{z})$, and $\mathbf{H}_{\rho'\alpha}(\mathbf{x},\mathbf{z})$ are calculated with x-z component data, where $\mathbf{H}_{\beta\alpha}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\rho'\alpha}(\mathbf{x},\mathbf{z})$ describe the mappings from α to β and ρ'. Then, spike model perturbations $\Delta\mathbf{m}_\beta(\mathbf{z}) = 0.1$ km/s and $\Delta\mathbf{m}_{\rho'}(\mathbf{z}) = 0.1$ g/cm^3 are applied respectively. MPSFs $\mathbf{H}_{\alpha\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\alpha\rho'}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\rho'}(\mathbf{x},\mathbf{z})$, and $\mathbf{H}_{\rho'\rho'}(\mathbf{x},\mathbf{z})$ are obtained. $\mathbf{H}_{\alpha\beta}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z})$ describe the mappings from β to α and ρ'. $\mathbf{H}_{\alpha\rho'}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\beta\rho'}(\mathbf{x},\mathbf{z})$ describe the mappings from ρ' to α and β.

These MPSFs are plotted in model space and arranged in a block structure in consistent with their positions in multiparameter Hessian, as shown in Figure 2a. Figure 2a is also equivalent to a sparse representation of multiparameter Hessian with 3 columns, which measure finite-frequency features of the interparameter tradeoffs. A positive α perturbation

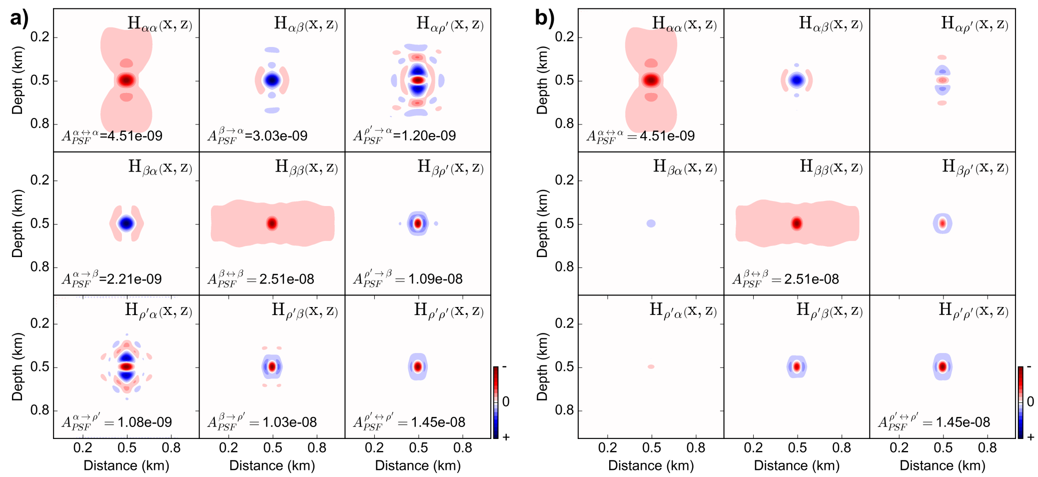

FIG. 2. Multiparameter point spread functions (MPSFs) of isotropic-elastic parameters in velocity-density parameterization. (a) shows the MPSFs $\mathbf{H}_{\alpha\alpha}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\alpha}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\rho'\alpha}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\alpha\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\alpha\rho'}(\mathbf{x},\mathbf{z})$, $\mathbf{H}_{\beta\rho'}(\mathbf{x},\mathbf{z})$, and $\mathbf{H}_{\rho'\rho'}(\mathbf{x},\mathbf{z})$ with x-z component data; (b) shows the corresponding normalized MPSFs. A_{PSF} indicate the maximum magnitudes of the MPSFs.

produces a negative contamination in β described by $\mathbf{H}_{\beta\alpha}(\mathbf{x},\mathbf{z})$ and vice versa. However, both positive α and β perturbations result in positive contaminations in density ρ' described by $\mathbf{H}_{\rho'\alpha}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z})$ and vice versa. Furthermore, regards to spatial spreading, the MPSFs representing contaminations to density (i.e., $\mathbf{H}_{\rho'\alpha}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\rho'\beta}(\mathbf{x},\mathbf{z})$) experience oscillatory side-lobes, which may distort the correct density updates.

As indicated by A_{PSF} in Figure 2a, magnitudes of the MPSFs, which describe relative strengths of the eigenvalues, also differ significantly. Magnitude of $\mathbf{H}_{\beta\beta}(\mathbf{x},\mathbf{z})$ is larger than the magnitudes of $\mathbf{H}_{\alpha\alpha}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\rho'\rho'}(\mathbf{x},\mathbf{z})$ meaning that the eigenvectors associated with S-wave velocity will be better recovered than those associated with P-wave velocity and density. To evaluate relative strengths of the interparameter contaminations, the contaminations are normalized with the MPSFs representing correct model updates. For example, the MPSFs $\mathbf{H}_{\alpha\beta}(\mathbf{x},\mathbf{z})$ and $\mathbf{H}_{\alpha\rho'}(\mathbf{x},\mathbf{z})$ are normalized by the maximum absolute value of $\mathbf{H}_{\alpha\alpha}(\mathbf{x},\mathbf{z})$. Normalized MPSFs are shown in Figure 2b. Contaminations from α to β and ρ' appear to be relatively weak. Density ρ' perturbations also produce moderate unwanted artifacts in α and β. S-wave velocity β suffers from the least amount of contaminations but produces strong mappings to α and ρ', which may make density under- or overestimated and cancel the updates for P-wave velocity. Geological features in the inverted P-wave velocity and density models may be contaminations from the S-wave velocity, which increases the uncertainties of the inverse problems significantly. These information helps us understand how the interparameter tradeoffs affect the inversion process.

To verify our analysis and predictions with MPSFs, inversion experiments with a Gaussian-anomaly model is carried out. Figures 3a, 3b, and 3c show the true P-wave velocity, S-wave velocity and density models with 3 isolated Gaussian anomalies. The initial models are homogeneous with $\alpha = 2.0$ km/s, $\beta = 1.2$ km/s and $\rho' = 1.2$ g/cm^3. The acquisition arrangement is the same with previous example. A l-BFGS optimization method is employed for updating α, β and ρ' simultaneously. This inversion experiment can be considered as an extended version of spike probing test with 3 Gaussian model perturbation vectors. Relative

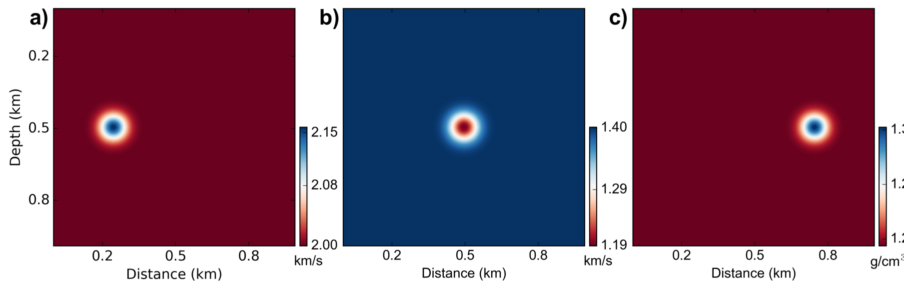

FIG. 3. Figures (a-c) show the true P-wave velocity, S-wave velocity and density of the Gaussian-anomaly model: $\mathbf{m}_{\alpha}^{\text{true}}$, $\mathbf{m}_{\beta}^{\text{true}}$ and $\mathbf{m}_{\rho'}^{\text{true}}$.

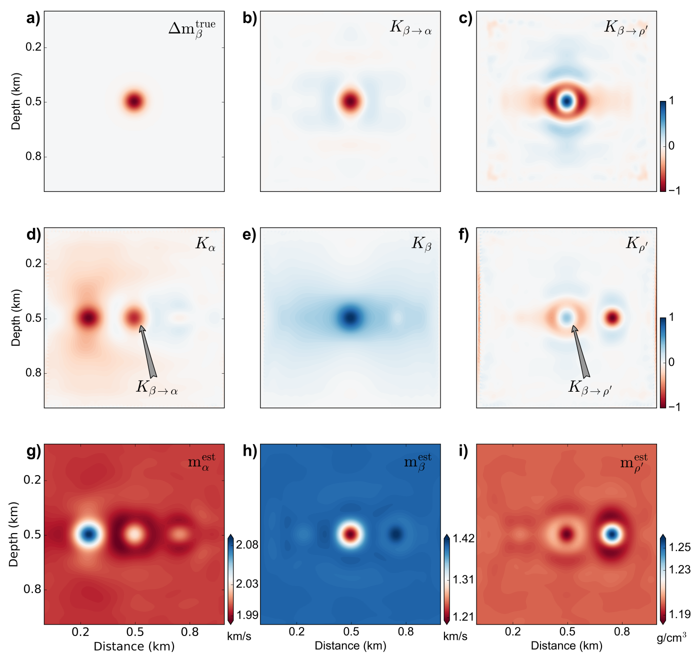

FIG. 4. Figures (a-c) show the true S-wave velocity model perturbation vector $\Delta\mathbf{m}_{\beta}^{\text{true}}$, true contamination kernels $K_{\beta\rightarrow\alpha}$ and $K_{\beta\rightarrow\rho'}$ respectively; Figures (d-f) illustrate the standard sensitivity kernels K_{α}, K_{β}, and $K_{\rho'}$; Figures (g-i) are the inverted models $\mathbf{m}_{\alpha}^{\text{est}}$ ($\tilde{\epsilon}_{\alpha}$=0.47), $\mathbf{m}_{\beta}^{\text{est}}$ ($\tilde{\epsilon}_{\beta}$=0.15) and $\mathbf{m}_{\rho'}^{\text{est}}$ ($\tilde{\epsilon}_{\rho'}$=0.77) after 10 iterations with traditional simultaneous inversion strategy.

least-squares error (RLSE) (equation (**??**)) is used to evaluate the quality of the inverted model. Figure 4a shows the true S-wave velocity model perturbation $\Delta\mathbf{m}_{\beta}^{\text{true}}$. Figures 4b and 4c show the true contamination kernels $K_{\beta\rightarrow\alpha}$ and $K_{\beta\rightarrow\rho'}$ calculated by multiplying multiparameter Hessian off-diagonal blocks $\mathbf{H}_{\alpha\beta}$ and $\mathbf{H}_{\rho'\beta}$ with true model perturbation vector $\Delta\mathbf{m}_{\beta}^{\text{true}}$. Figures 4d, 4e and 4f show the standard sensitivity kernels K_{α}, K_{β} and $K_{\rho'}$ in the first iteration. Strengths and characteristics of the interparameter contaminations generally match our predictions using MPSFs shown in Figure 2. A negative S-wave velocity perturbation produces strong positive and negative contaminations into the updates for α and ρ', as indicated by the *interparameter contamination kernels* $K_{\beta\rightarrow\alpha}$ and $K_{\beta\rightarrow\rho'}$. Figures 4g, 4h and 4i show the inverted P-wave velocity, S-wave velocity and density models after 10 iterations using traditional simultaneous inversion strategy. S-wave velocity suffers from limited contaminations and is best inverted. P-wave velocity and density suffer strong contaminations from S-wave velocity. As iteration proceeds, S-wave velocity

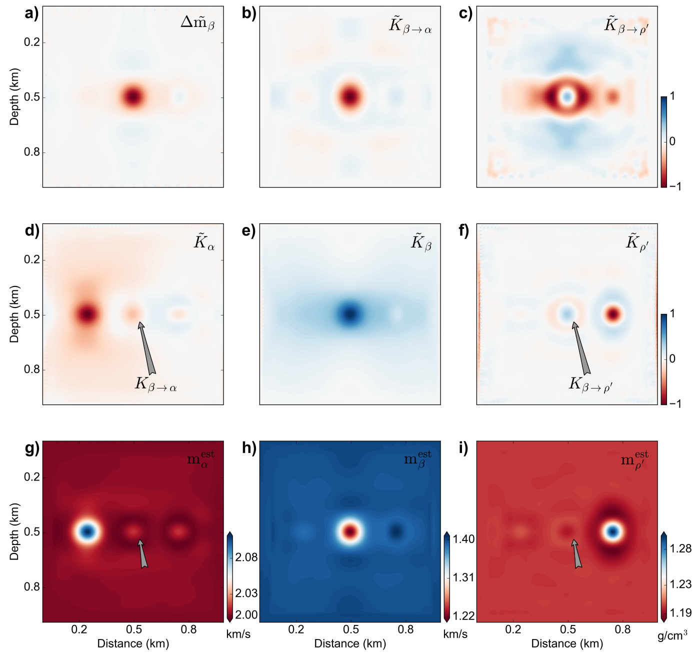

FIG. 5. Figures (a-c) show the estimated S-wave velocity model perturbation vector $\Delta\tilde{\mathbf{m}}_\beta$ ($k' = 3$), approximate contamination kernels $\tilde{K}_{\beta\to\alpha}$ and $\tilde{K}_{\beta\to\rho'}$ respectively; Figures (d-f) illustrate the new update kernels $\tilde{K}_\alpha$, $\tilde{K}_\beta$, and $\tilde{K}_{\rho'}$; Figures (g-i) are the inverted models $\mathbf{m}_\alpha^{\text{est}}$ ($\tilde{\epsilon}_\alpha$=0.32), $\mathbf{m}_\beta^{\text{est}}$ ($\tilde{\epsilon}_\beta$=0.14) and $\mathbf{m}_{\rho'}^{\text{est}}$ ($\tilde{\epsilon}_{\rho'}$=0.61) after 10 iterations with new inversion strategy.

is estimated fastest. The interparameter contaminations due to S-wave velocity perturbations are also reduced iteratively and if a sufficient number of iterations are performed, the contaminations are expected to be removed almost completely.

Figure 5a shows the estimated S-wave velocity model perturbation vector $\Delta\mathbf{m}_\beta^{\text{est}}$ after $k' = 3$ iterations. Figures 5b and 5c show the approximate contamination kernels $\tilde{K}_{\beta\to\alpha}$ and $\tilde{K}_{\beta\to\rho'}$ calculated by multiplying multiparameter Hessian off-diagonal blocks $\mathbf{H}_{\alpha\beta}$ and $\mathbf{H}_{\rho'\beta}$ with estimated model perturbation vector $\Delta\mathbf{m}_\beta^{\text{est}}$. The features of the approximate contamination kernels match those of true contamination kernels (Figures 4b and 4c) very well. Figures 5d, 5e and 5f are the new update kernels following equation (55). Figures 5g, 5h and 5i show the inverted P-wave velocity, S-wave velocity and density models with the new inversion strategy. As indicated by the arrows, contaminations from S-wave velocity to P-wave velocity and density have been suppressed. Figure 6 shows the convergence histories of traditional simultaneous inversion strategy and new inversion strategy for the Gauss-anomaly example. The new inversion strategy provides faster convergence compared to traditional simultaneous inversion strategy but it is 2.5 times more expensive.

Marmousi model example

The proposed stochastic probing strategy is first applied to evaluate the strengths of the interparameter tradeoffs within the whole volume. The new inversion strategy with approximate contamination kernels is employed to invert the isotropic-elastic parameters in comparison with traditional simultaneous inversion strategy. Approximate eigenvalue vol-

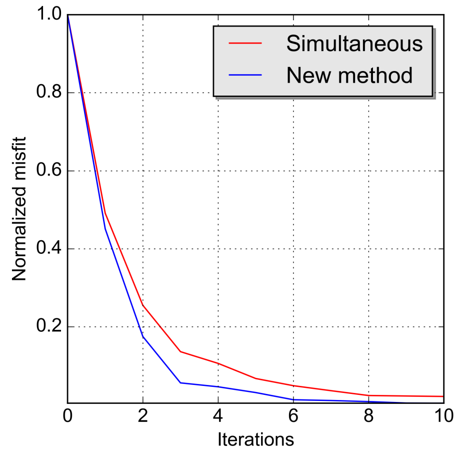

FIG. 6. Convergence histories comparison of traditional simultaneous inversion strategy (red curve) and new inversion strategy (blue curve) for the Gaussian-anomaly inversion example.

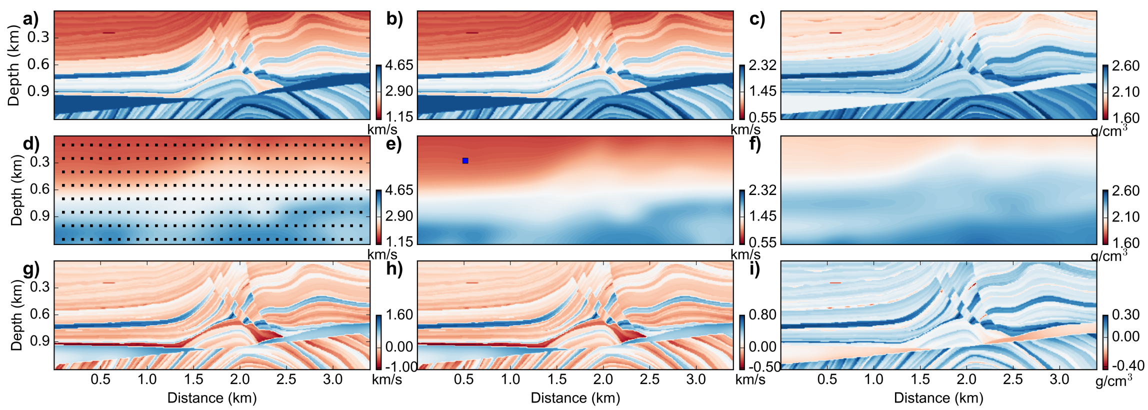

FIG. 7. (a-c) show true P-wave velocity, S-wave velocity and density models; (d-f) show initial P-wave velocity, S-wave velocity and density models; Figures (g-i) show true P-wave velocity, S-wave velocity and density model perturbations. The regularly distributed black squares in (d) represent the vector $\mathbf{v}'$ for interparameter tradeoffs analysis within the whole model. The blue square in (e) indicates the location $\mathbf{z}_1 = (0.515\ \mathrm{km}, 0.275\ \mathrm{km})$ for quantifying local spatial and interparameter tradeoffs of the inverted models.

umes and extended multiparameter point spread functions (EMPSFs) are used to evaluate resolution of the inverted models.

Figures 7a, 7b and 7c show the true P-wave velocity, S-wave velocity and density models. Figures 7d, 7e and 7f show the initial P-wave velocity, S-wave velocity and density models. Figure 7g, 7h and 7i show the corresponding true model perturbations. The model is 3.4 km wide and 1.2 km deep. Number of 33 sources and 330 receivers are deployed regularly with a source spacing of 100 m and a receiver spacing of 10 m along top surface of the model. A Ricker wavelet with dominant frequency of 6 Hz is used for forward modelling.

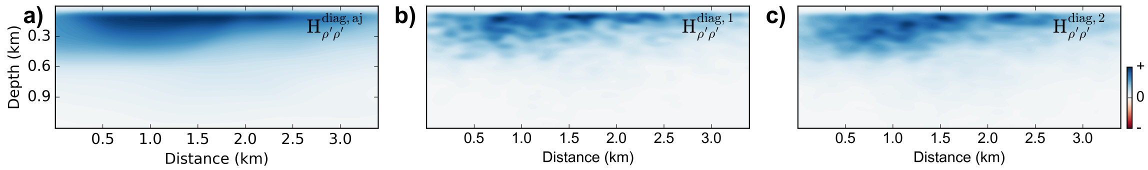

FIG. 8. (a) shows the Hessian diagonals $\mathbf{H}^{\text{diag, aj}}_{\rho'\rho'}$ calculated with adjoint-state method; (b-c) show the stochastic estimation of Hessian diagonals $\mathbf{H}^{\text{diag, }1}_{\rho'\rho'}$ and $\mathbf{H}^{\text{diag, }2}_{\rho'\rho'}$ with 1 and 2 random vector applications respectively.

Evaluating the strengths of interparameter tradeoffs within the whole volume

Diagonals of subblock matrices in multiparameter Hessian are first estimated with the stochastic probing approach following equation (47). Figure 8a shows the Hessian diagonals $\mathbf{H}^{\text{diag, aj}}_{\rho'\rho'}$ calculated with adjoint-state method (Shin et al., 2001). The computation cost is equivalent to 363 forward simulations. Figures 8b and 8c show the stochastic estimations of Hessian diagonals with 1 and 2 random vector applications respectively. Computation costs are equivalent to 66 and 198 forward simulations. Energy distributions in the stochastic estimations generally match those calculated with the adjoint-state method, which verifies the effectiveness of stochastic probing approach. Figures 9a show the stochastic estimations of the Hessian diagonals $\mathbf{H}^{\text{diag}}_{\alpha\alpha}$, $\mathbf{H}^{\text{diag}}_{\beta\beta}$ and $\mathbf{H}^{\text{diag}}_{\rho'\rho'}$ after normalization. Energy distributions in these Hessian diagonals for different parameters differ significantly. Stronger elements of the Hessian diagonals mean that the model parameters are well constrained. However, energies of $\mathbf{H}^{\text{diag}}_{\beta\beta}$ are constrained in the shallow parts of the model. Maximum magnitudes of $\mathbf{H}^{\text{diag}}_{\beta\beta}$ are approximately 11.0 times and 6.3 times stronger than those of $\mathbf{H}^{\text{diag}}_{\alpha\alpha}$ and $\mathbf{H}^{\text{diag}}_{\rho'\rho'}$, which means that S-wave velocity will be better recovered than P-wave velocity and density. The Hessian diagonals are also used as preconditioners in the inversion process.

Figures 9b show the stochastic estimations of the Hessian diagonals $\mathbf{H}^{\text{diag}}_{\alpha\beta}$, $\mathbf{H}^{\text{diag}}_{\alpha\rho'}$ and $\mathbf{H}^{\text{diag}}_{\beta\rho'}$ which measure the coupling strengths of the isotropic-elastic parameters in the whole volume. The coupling strengths change within the whole volume significantly, that is to say, they are influenced by inhomogeneity of the model and source-receive illumination. In earlier iterations, strong interparameter tradeoffs appear at the shallow parts of the model, as indicated by the grey and white arrows. Magnitudes of the diagonals of off-diagonal blocks associated with different physical parameters are quite different. $\mathbf{H}^{\text{diag}}_{\beta\rho'}$ is much stronger than $\mathbf{H}^{\text{diag}}_{\alpha\beta}$ and $\mathbf{H}^{\text{diag}}_{\alpha\rho'}$ meaning that the interparameter tradeoffs among the isotropic-elastic parameters mainly come from the coupling effects between S-wave velocity and density. $\mathbf{H}^{\text{diag}}_{\beta\rho'}$ is very similar to $\mathbf{H}^{\text{diag}}_{\beta\beta}$ meaning that the coupling effects between S-wave velocity and density are dominated by S-wave velocity.

In this research, a vector $\mathbf{v}'$, which consists of regularly distributed spikes with a constant magnitude of 0.2, is designed, as indicated by the black squares in Figure 7d. Products of the multiparameter Hessian subblocks approximate row summations of the multiparameter Hessian, as illustrated in Figure 10. Strengths of interparameter contaminations generally match our predictions with multiparameter Hessian diagonals. Areas with strong interparameter tradeoffs are indicated by the grey and white arrows in Figure 10. Fur-

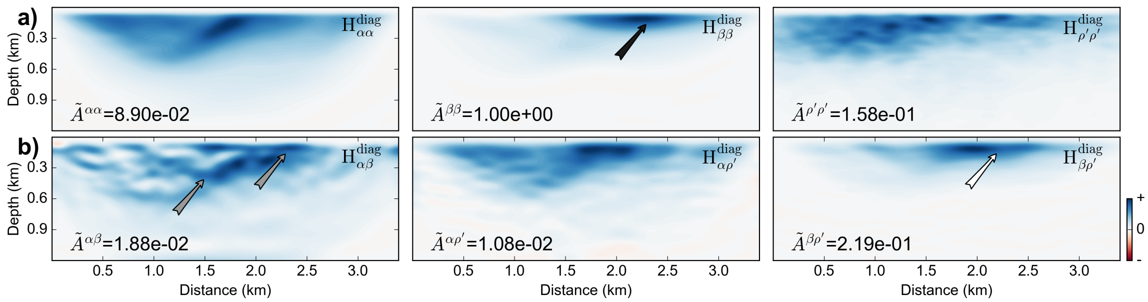

FIG. 9. (a) shows the the stochastic estimations of Hessian diagonals $\mathbf{H}^{\text{diag}}_{\alpha\alpha}$, $\mathbf{H}^{\text{diag}}_{\beta\beta}$ and $\mathbf{H}^{\text{diag}}_{\rho'\rho'}$ with 2 random vector applications; (b) shows the stochastic estimations of Hessian diagonals $\mathbf{H}^{\text{diag}}_{\alpha\beta}$, $\mathbf{H}^{\text{diag}}_{\alpha\rho'}$ and $\mathbf{H}^{\text{diag}}_{\beta\rho'}$ with 2 random vector applications. $\tilde{A}$ mean the maximum magnitude of the Hessian diagonals after normalization.

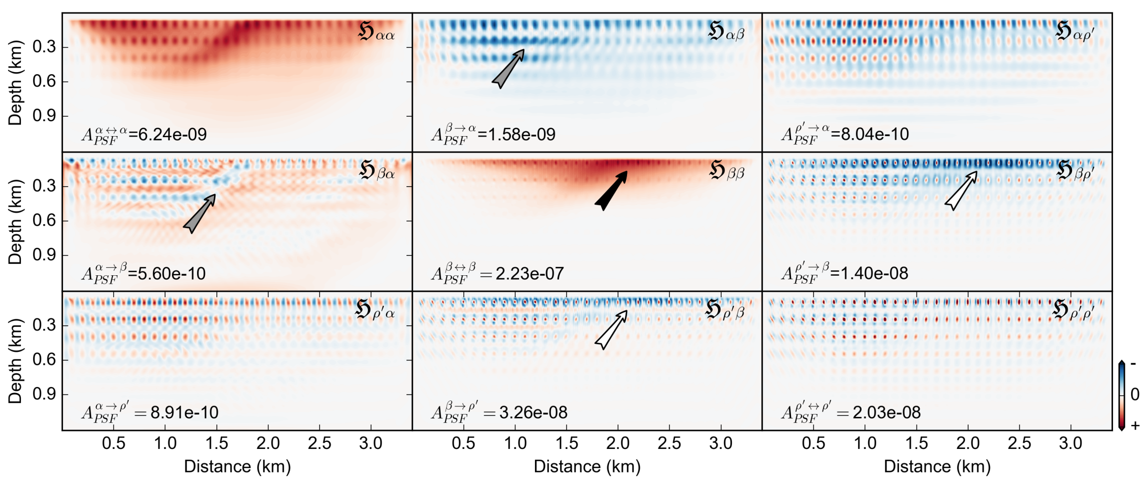

FIG. 10. Products of multiparmaeter Hessian with vector $\mathbf{v}'$. The first column show the multiparameter Hessian-vector products $\mathfrak{H}_{\alpha\alpha} = \mathbf{H}_{\alpha\alpha}\mathbf{v}'$, $\mathfrak{H}_{\beta\alpha} = \mathbf{H}_{\beta\alpha}\mathbf{v}'$, and $\mathfrak{H}_{\beta\alpha} = \mathbf{H}_{\beta\alpha}\mathbf{v}'$. The second column show the multiparameter Hessian-vector products $\mathfrak{H}_{\alpha\beta} = \mathbf{H}_{\alpha\beta}\mathbf{v}'$, $\mathfrak{H}_{\beta\beta} = \mathbf{H}_{\beta\beta}\mathbf{v}'$, and $\mathfrak{H}_{\rho'\beta} = \mathbf{H}_{\rho'\beta}\mathbf{v}'$. The third column show the multiparameter Hessian-vector products $\mathfrak{H}_{\alpha\rho'} = \mathbf{H}_{\alpha\rho'}\mathbf{v}'$, $\mathfrak{H}_{\beta\rho'} = \mathbf{H}_{\beta\rho'}\mathbf{v}'$, and $\mathfrak{H}_{\rho'\rho'} = \mathbf{H}_{\rho'\rho'}\mathbf{v}'$.

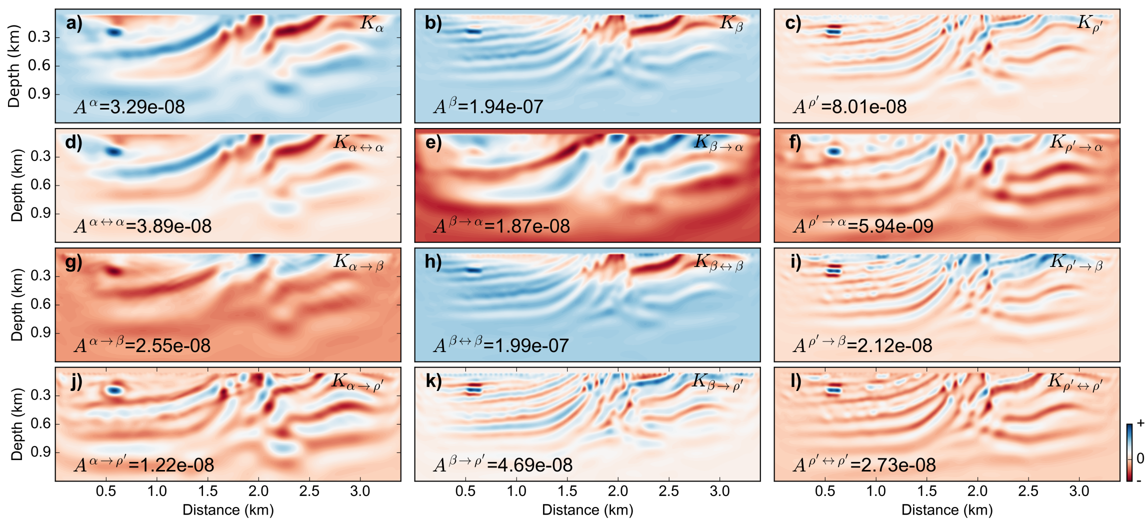

FIG. 11. (a-c) illustrate the standard sensitivity kernels K_α, K_β, and $K_{\rho'}$; (d-f) show the correct update kernel $K_{\alpha\leftrightarrow\alpha}$ and contamination kernels $K_{\beta\to\alpha}$ and $K_{\rho'\to\alpha}$; (g-i) show the contamination kernel $K_{\alpha\to\beta}$, correct update kernels $K_{\beta\leftrightarrow\beta}$ and contamination kernel $K_{\rho'\to\beta}$; (j-l) show contamination kernels $K_{\beta\to\rho'}$ and $K_{\alpha\to\rho'}$ and correct update kernel $K_{\rho'\leftrightarrow\rho'}$. A represent maximum magnitudes of the kernels.

thermore, comparing strengths of the off-diagonal Hessian-vector products (i.e., $\mathfrak{H}_{\beta\alpha}$) with those of diagonal Hessian-vector products (i.e., $\mathfrak{H}_{\beta\beta}$), it is concluded that the contaminations from α to β and ρ' are relatively weak and can be ignored. Contaminations from ρ' to α and β are also not very strong. However, the contaminations from β to α may degrade the update for α. Contaminations from β to ρ' may boost the density update by 1.8 times.

To verify these predictions and conclusions, the true *interparameter contamination kernels* are calculated by applying multiparameter Hessian to the true model perturbation vectors $\Delta\mathbf{m}_\alpha$, $\Delta\mathbf{m}_\beta$ and $\Delta\mathbf{m}_{\rho'}$ as shown in Figures 7g, 7h and 7i. The first row in Figure 11 show the standard sensitivity kernels K_α, K_β and $K_{\rho'}$, which are contaminated by mappings from other parameters. The second row in Figure 11 show the correct update kernel $K_{\alpha\leftrightarrow\alpha}$ and contamination kernels $K_{\beta\to\alpha}$ and $K_{\rho'\to\alpha}$. In the third row of Figure 11, the contamination kernel $K_{\alpha\to\beta}$, correct update kernel $K_{\beta\leftrightarrow\beta}$, and $K_{\rho'\to\beta}$ are illustrated from left to right. In the forth row of Figure 11, contamination kernels $K_{\alpha\to\rho'}$ and $K_{\beta\to\rho'}$ and correct update kernel $K_{\rho'\to\rho'}$ are given.

Since magnitudes of the true model perturbation vectors change within the whole volume and the strengths of P-wave velocity perturbation are approximately 2 times and 4 times larger than those of S-wave velocity and density perturbations, the contamination kernels are not entirely consistent with the predictions by Hessian diagonals and Hessian-vector products shown in Figures 9 and 10 exactly. I interpret this is an indication of the complexity of the resolution problem in general. However, areas with strong elements in $K_{\beta\leftrightarrow\beta}$ generally match those of Hessian diagonals $\mathbf{H}^{\text{diag}}_{\beta\beta}$ (Figure 9) and Hessian-vector product $\mathfrak{H}_{\beta\beta}$ (Figure 10), as indicated by the black arrows. Examining the contamination kernels $K_{\alpha\to\beta}$ and $K_{\beta\to\alpha}$ tells us that Hessian diagonals ($\mathbf{H}^{\text{diag}}_{\alpha\beta}$ in Figure 9) and Hessian-vector products ($\mathfrak{H}_{\alpha\beta}$ and $\mathfrak{H}_{\beta\alpha}$ in Figure 9) predict energy distributions of the interparameter tradeoffs, as indicated by the grey arrows. White arrows in $K_{\rho'\to\beta}$ and $K_{\beta\to\rho'}$ also

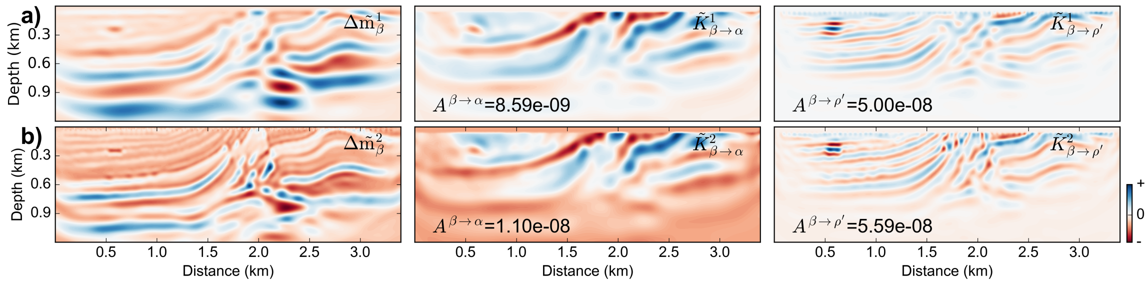

FIG. 12. (a) shows the estimated model perturbation vector $\Delta\tilde{\mathbf{m}}^1_\beta$, approximate contamination kernels $\tilde{K}^1_{\beta\to\alpha}$ and $\tilde{K}^1_{\beta\to\rho'}$; (b) shows the estimated model perturbation vector $\Delta\tilde{\mathbf{m}}^2_\beta$, approximate contamination kernels $\tilde{K}^2_{\beta\to\alpha}$ and $\tilde{K}^2_{\beta\to\rho'}$.

indicate the areas with strong interparameter tradeoffs between S-wave velocity and density.

Comparing magnitudes of the correct updates and *interparameter contamination kernels*, it can be observed that $K_{\beta\to\beta}$ is very close to K_β meaning that the S-wave velocity suffers limited contamination from α and ρ'. Furthermore, the correct update kernel $K_{\alpha\leftrightarrow\alpha}$ will be degraded by the contamination kernel $K_{\beta\to\alpha}$. Contamination kernel $K_{\beta\to\rho'}$ is approximately 1.7 times stronger than the correct update kernel $K_{\rho'\leftrightarrow\rho'}$, which will make density highly under- or overestimated. The contaminations from β to ρ' will dominate the estimated density structures. Note: during the inversion process, the contaminations can be reduced partially and the energy distributions of the interparameter contaminations may also change.

Mitigating the interparameter tradeoffs

To mitigate the contamination of S-wave velocity into other parameters, a novel inversion strategy is proposed with approximate contamination kernels. I first carry out inversion experiments by $k' = 8$ and 15 iterations, which provide estimated model perturbation vectors $\Delta\tilde{\mathbf{m}}^1_\beta$ and $\Delta\tilde{\mathbf{m}}^2_\beta$, as shown in Figures 12a and 12b. The estimated P-wave and density perturbations are dropped. Contamination kernels $\tilde{K}^1_{\beta\to\alpha}$, $\tilde{K}^1_{\beta\to\rho'}$, $\tilde{K}^2_{\beta\to\alpha}$, and $\tilde{K}^2_{\beta\to\rho'}$ are constructed by applying multiparameter Hessian off-diagonal blocks $\mathbf{H}_{\alpha\beta}$ and $\mathbf{H}_{\rho'\beta}$ to the estimated model vectors, as shown in Figures 12a and 12b. Magnitudes and characteristics of the approximate contamination kernels match the true contamination kernels $\tilde{K}_{\beta\to\alpha}$ and $\tilde{K}_{\beta\to\rho'}$ shown in Figure 11 very well. Because $\Delta\tilde{\mathbf{m}}^2_\beta$ is more resolved and better recovered than $\Delta\tilde{\mathbf{m}}^1_\beta$, $\tilde{K}^2_{\beta\to\alpha}$ and $\tilde{K}^2_{\beta\to\rho'}$ represent better approximations than $\tilde{K}^1_{\beta\to\alpha}$ and $\tilde{K}^1_{\beta\to\rho'}$.

The new updates kernels $\tilde{K}^1_\alpha$, $\tilde{K}^1_\beta$, $\tilde{K}^1_{\rho'}$, $\tilde{K}^2_\alpha$, $\tilde{K}^2_\beta$ and $\tilde{K}^2_{\rho'}$ are calculated by subtracting the approximate contamination kernels from the standard sensitivity kernels following equation (55), as shown in Figure 13. Magnitudes of the new updates kernels $\tilde{K}^1_{\rho'}$ and $\tilde{K}^2_{\rho'}$ have been reduced by approximately 38.2% and 56.8%. In particular, it is observed that the features of new update kernel $\tilde{K}^2_{\rho'}$ for density are very close to the characteristics of true update kernel $K_{\rho'\leftrightarrow\rho'}$ shown in Figure 11, which means that the contaminations from S-wave velocity to density have been suppressed.

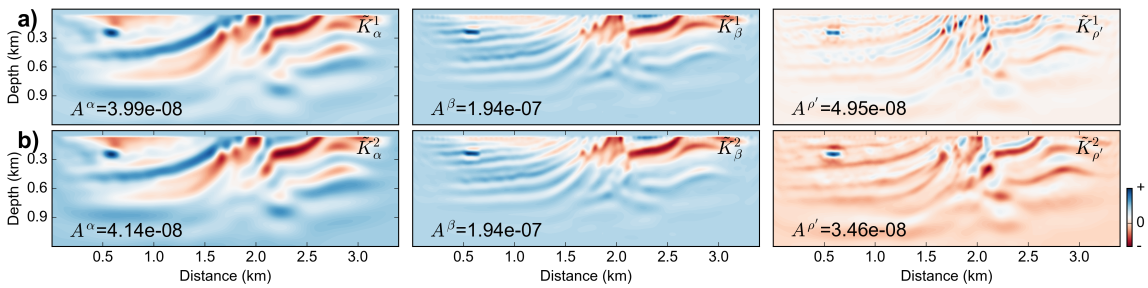

FIG. 13. (a) show the new update kernels $\tilde{K}_\alpha^1$, $\tilde{K}_\beta^1$ and $\tilde{K}_{\rho'}^1$; (b) show the new update kernels $\tilde{K}_\alpha^2$, $\tilde{K}_\beta^2$ and $\tilde{K}_{\rho'}^2$.

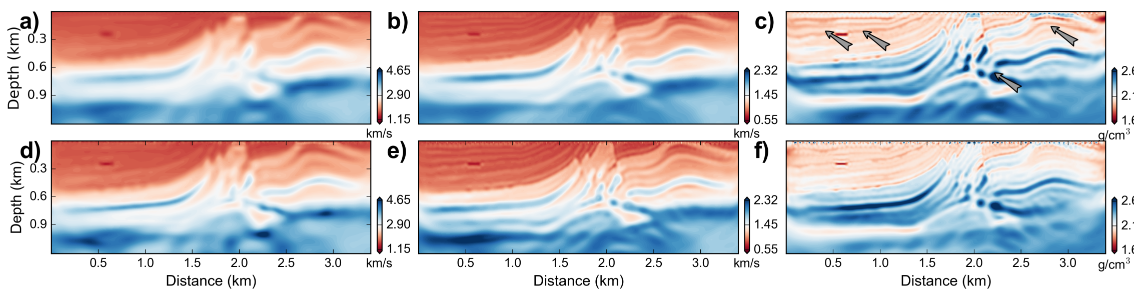

FIG. 14. (a-c) show inverted P-wave velocity ($\tilde{\epsilon}_\alpha$=0.83), S-wave velocity ($\tilde{\epsilon}_\beta$=0.72) and density ($\tilde{\epsilon}_{\rho'}$=1.04) models with traditional simultaneous inversion strategy; (d-f) show the inverted P-wave velocity ($\tilde{\epsilon}_\alpha$=0.76), S-wave velocity ($\tilde{\epsilon}_\beta$=0.67) and density ($\tilde{\epsilon}_{\rho'}$=0.83) models using new inversion method with approximate contamination kernels.

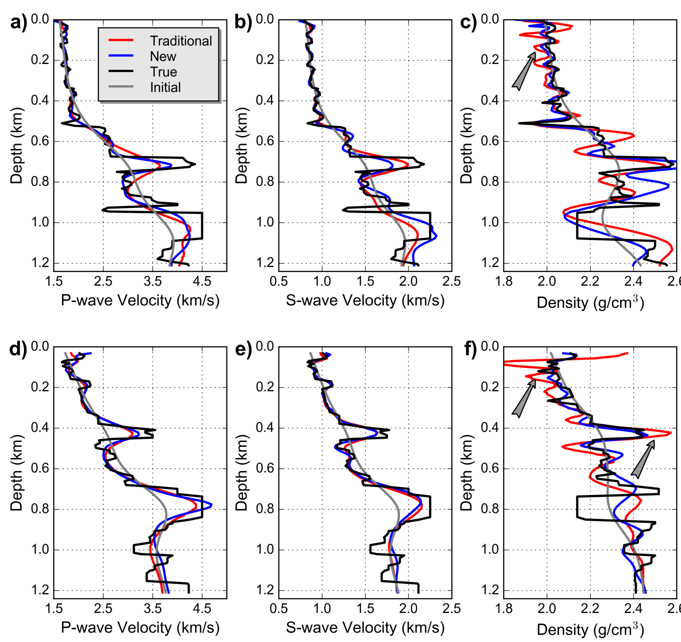

FIG. 15. (a-c) show the well log data of P-wave velocity, S-wave velocity, and density models at 0.5 km; (d-f) show the well log data at 3.0 km. The red and grey curves indicate the true and initial models. The blue and green lines indicate the inverted models by traditional simultaneous inversion strategy and new inversion strategy.

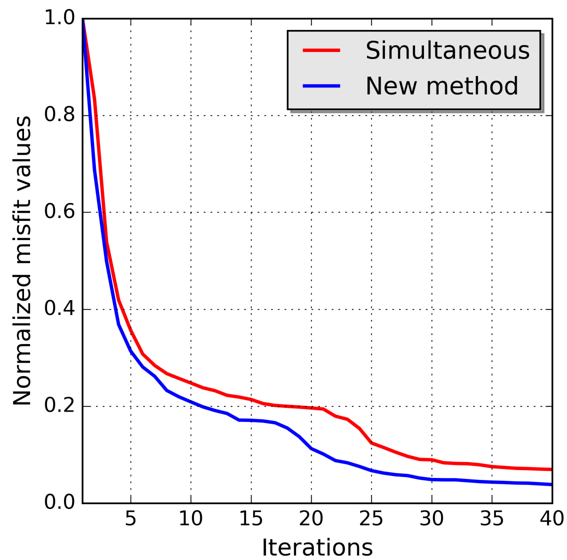

FIG. 16. Convergence history comparison of traditional simultaneous inversion strategy (red curve) and new inversion strategy (blue curve) for Marmousi model example.

The P-wave velocity α, S-wave velocity β and density ρ' are inverted with traditional simultaneous inversion strategy. Multiscale approach is adopted for reducing the nonlinearity by expanding the frequency band from [3 Hz, 5Hz] to [3 Hz, 8Hz]. With each frequency band, 20 iterations are performed. Figures 14a, 14b and 14c show the inverted α, β and ρ' models after 40 iterations with traditional simultaneous inversion strategy. Following the work-flow illustrated in Table 4.1, I first simultaneously update α, β and ρ' models by $k' = 15$ iterations and the inverted models are dropped but only keep the estimated model perturbation $\Delta\tilde{\mathbf{m}}_\beta$. Then, model parameters α, β and ρ' are inverted again from initial models by 10 iterations. In this inversion loop, at each iteration approximate contamination kernels are constructed and the models are updated with new kernels as indicated in equation (55). This process is then repeated every 10 iterations. Figures 14d, 14e and 14f show the inverted P-wave velocity, S-wave velocity and density models with the new inversion strategy after 40 iterations. The computation cost is 2.5 times more expensive than traditional simultaneous inversion strategy.

S-wave velocity is best inverted and more resolved than P-wave velocity. P-wave velocity is poorly recovered but limited interparameter contamination artifacts can be observed. With traditional simultaneous inversion strategy, the S-wave velocity structures are mapped into the estimated density model as indicated by the arrows in Figures 14c, 15c and 15f. Positive S-wave velocity perturbations make density overestimated and negative S-wave velocity perturbations make density underestimated. With the new inversion strategy, the imprints in the inverted density model have been suppressed effectively, as indicated by arrows in Figures 14f, 15c and 15f. Furthermore, the inverted P-wave velocity (Figure 14d) and S-wave velocity (Figure 14e) are also enhanced. The new inversion approach is also able to provide faster convergence, as shown in Figure 16.

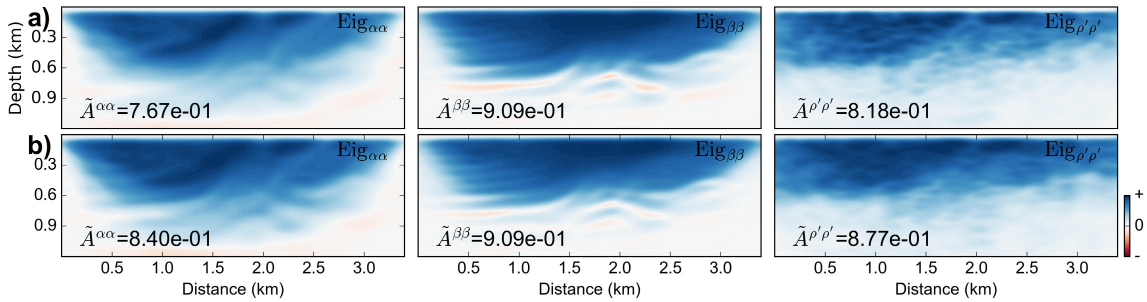

FIG. 17. Approximate eigenvalue volumes of the inverted models. Figures (a-c) show approximate eigenvalue volumes $\mathbf{Eig}_{\alpha\alpha}$, $\mathbf{Eig}_{\beta\beta}$, and $\mathbf{Eig}_{\rho'\rho'}$ for inverted P-wave velocity, S-wave velocity and density by traditional simultaneous inversion strategy; Figures (d-f) show the corresponding approximate eigenvalue volumes by the new inversion strategy.

Resolution analysis

Approximate eigenvalue volume (equation (64)) and extended multiparameter point spread functions (EMPSFs) are used to quantify resolution of the inverted models with different inversion strategies. Figures 17a, 17b and 17c show the approximate eigenvalue volumes obtained with 2 random Hessian-vector applications. The approximate eigenvalue volumes of S-wave velocity ($\mathbf{Eig}_{\beta\beta}$) are closer to 1 than those of P-wave velocity and density, which means that S-wave velocity is better recovered than P-wave velocity and density. Magnitudes of the approximate eigenvalue volumes decrease with increasing depths meaning that shallow parts of models are better recovered than deep parts.

Spike model perturbations $\Delta\mathbf{m}_\alpha = 0.2$ km/s, $\Delta\mathbf{m}_\beta = 0.2$ km/s and $\Delta\mathbf{m}_{\rho'} = 0.2$ g/cm^3 are applied at local position $\mathbf{z}_1 = (0.515\,\text{km}, 0.275\,\text{km})$ (as shown in Figure 7e) respectively, which are used to measure the local spatial and interparameter tradeoffs of the inverted models. The traditional MPSFs are plotted in Figure 18a. The MPSFs representing interparameter contaminations are normalized with those representing correct model estimations. Strong contaminations from S-wave velocity to P-wave velocity and density are observed as indicated by the grey arrows. The spike model perturbations are then reconstructed with 10 conjugate-gradient iterations, which gives the EMPSFs, as presented in Figures 18b. Contaminations from S-wave velocity to P-wave velocity and density are reduced by approximately 23.8% and 47.1%, as indicated by the grey arrows. Furthermore, compared to MPSFs, EMPSFs are more de-blurred. This is strongly suggestive that the local spatial and interparameter tradeoffs provided by traditional MPSFs may not be accurate and the inverted models by traditional simultaneous inversion strategy suffer strong interparameter tradeoffs. In Figures 18c, the EMPSFs obtained with conjugate-gradient method following the work-flow with approximate contamination kernels are given. The contaminations from S-wave velocity to P-wave velocity and density are reduced effectively, as indicated by grey arrows. This means that the inverted models (Figures 14a, 14b and 14c) by new inversion strategy suffer little interparameter contaminations.

Hussar dataset application

At the end, the proposed strategies are applied to invert isotropic-elastic parameters with Hussar practical seismic dataset and quantify resolution of the inverted models. In Septem-

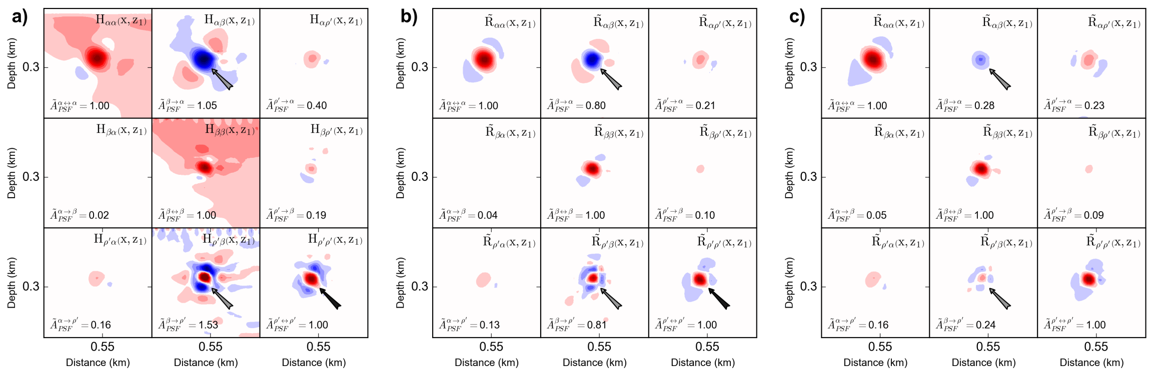

FIG. 18. (a) show the traditional MPSFs after normalization at position $\mathbf{z}_1$; (b) show the normalized EMPSFs with 10 conjugate-gradient iterations; (c) show the normalized EMPSFs constructed with new inversion work-flow (Table 4.1).

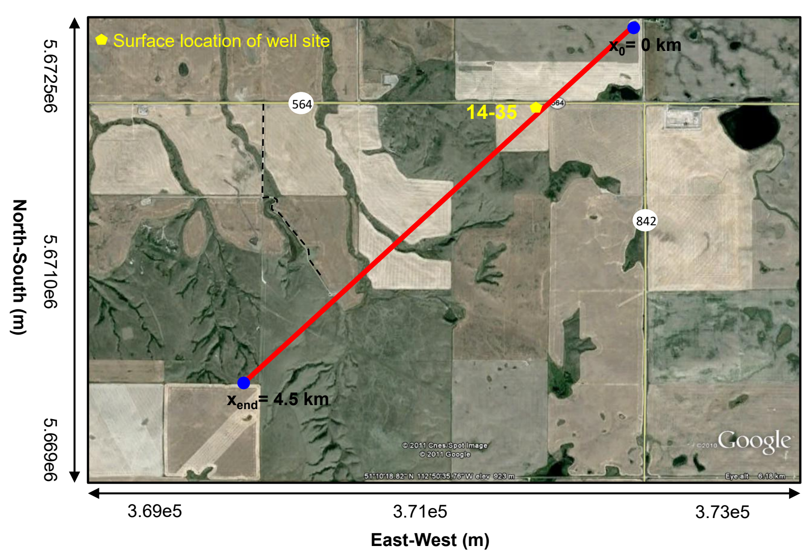

FIG. 19. The location of seismic line and well (14-35) in Hussar experiment (Margrave et al., 2012). Note: I have reset the coordinate of the seismic line for FWI. I assume that initial location of the seismic line starts at x_0=0 km and ends at x_{end}=4.5 km, as indicated by the blue circles.

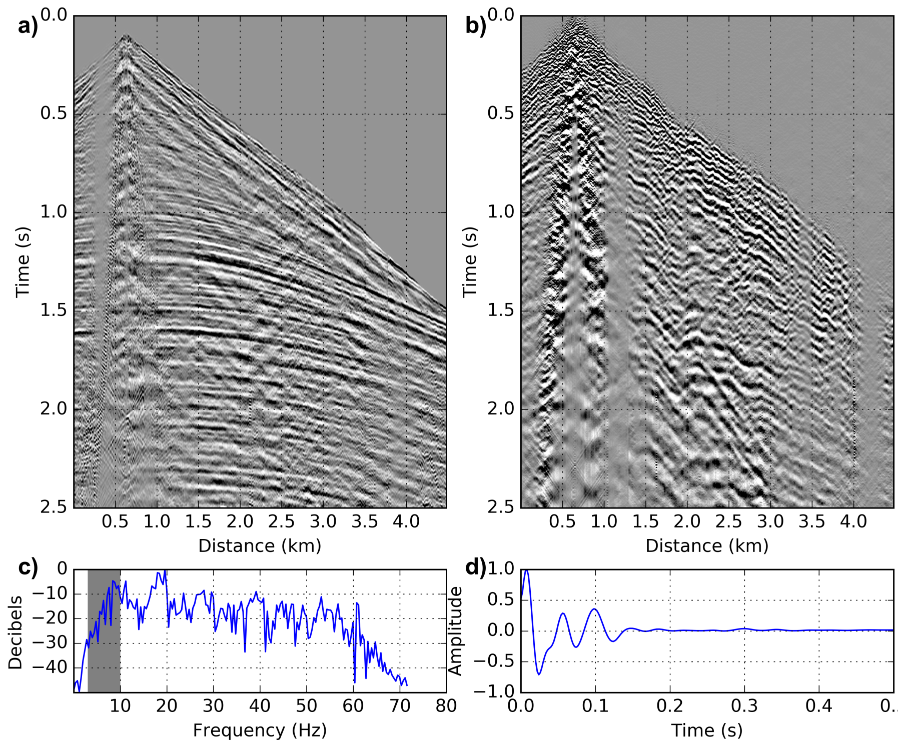

FIG. 20. (a) and (b) show the preprocessed vertical (z) and radial (x) component shot gathers at position of 0.6 km in horizontal distance; (c) shows the amplitude spectrum of the data. The shaded area means frequency band of [3Hz, 10Hz]. (d) shows the estimated minimum phase wavelet with dominant frequency of 25 Hz.

ber 2011, CREWES (Consortium for Research in Elastic Wave Exploration Seismology) initiated a seismic experiment in Hussar area, which is about 100 km east of Calgary, Alberta, Canada. The objective of this experiment was to maximize the low frequency content of the seismic data (Margrave et al., 2012), and to acquire a land dataset maximally suitable for full-waveform inversion methods. The 2D seismic survey line is 4.5 km in length. Figure 19 show the locations of the seismic line and well log 14-35. The seismic experiments were carried out with dynamite and vibroseis sources and different receiver types. In this research, I use the multicomponent data recorded by 10 Hz 3C (three-component) geophones with dynamite sources for inversion. A total of 269 sources (2 kg charge at 15 m in depth) are arranged regularly with a spacing of 20 m. A total of 448 geophones are distributed with a spacing of 10 m.

The raw seismic shot gathers are preprocessed with a series of steps. Automatic gain control (AGC) is first applied for amplitude recovery. Surface waves and monochromatic noise are suppressed with F-K filtering. Elevation statics and residual statics are applied to compensate the topographic variations and near-surface lithological variations. The seismic data is finally band-pass filtered within the frequency band of [3Hz, 60Hz]. Figures 20a and 20b show the preprocessed vertical (z) and radial (x) component data. Figure 20c shows the amplitude spectrum of the data. Frequency band of [3Hz, 10Hz] is used for inversion, as indicated by the shaded area. A minimum phase wavelet with dominant frequency of 25 Hz is estimated from seismic data and used for forward modelling, as illustrated in Figure 20d.

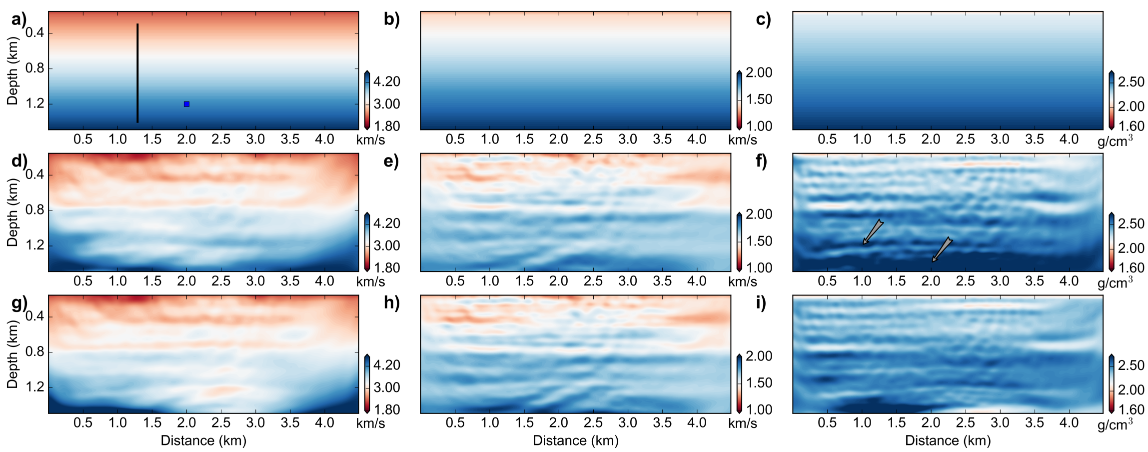

FIG. 21. (a-c) show the initial P-wave velocity, S-wave velocity and density models; (d-f) show the inverted P-wave velocity, S-wave velocity and density models using traditional simultaneous inversion strategy; (g-i) show the inverted P-wave velocity, S-wave velocity and density models using new inversion method. The black line in (a) indicates the position of well log $14\text{-}35$. The blue square in (a) indicates the location $\mathbf{z}_1$=$(2.0$ km$, 1.2$ km$)$ for local spatial and interparameter tradeoffs analysis.

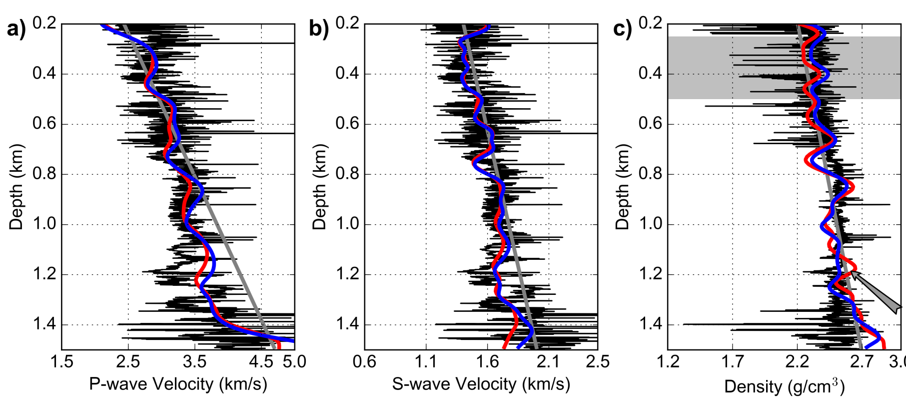

FIG. 22. (a-c) show the well log data comparison of P-wave velocity, S-wave velocity and density respectively. Black and gray curves are true well log data and initial models. Red and blue curves are the inverted models by traditional simultaneous inversion method and new inversion method respectively.

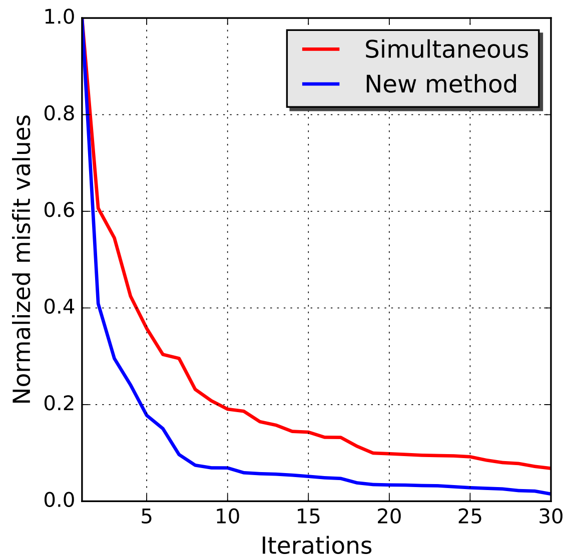

FIG. 23. Convergence history comparison of traditional simultaneous inversion method (red) and new inversion method (blue) for Hussar seismic dataset.

Figures 21a, 21b and 21c show the linear initial P-wave velocity, S-wave velocity and density models. The well log (14-35) is located at about 1.29 km in horizontal distance, as indicated by the black line in Figure 21a. I first simultaneously update P-wave velocity, S-wave velocity and density by expanding the frequency band from [3Hz, 5Hz] to [3Hz, 8Hz] and then [3Hz, 10Hz] with 10 iterations for each frequency band. The inverted P-wave velocity, S-wave velocity and density models using this traditional inversion strategy are illustrated in Figures 21d, 21e and 21f. I then carry out inversion experiments using new inversion strategy with approximate contamination kernels following the work-flow shown in Table 4.1. At each frequency band, I need to simultaneously update the model parameters by k'=15 iterations for estimating the approximate S-wave velocity perturbation vector $\Delta\tilde{\mathbf{m}}_\beta$. Figures 21g, 21h and 21i show the inverted P-wave velocity, S-wave velocity and density models using the new inversion method. Some artifacts appear in the inverted models. The geological layers are resolved and most of them are flat, which are consistent with the previous studies of impedance inversion (Lloyd, 2013; Cui, 2015; Esmaeili, 2016). Figures 22a, 22b and 22c show the well log data comparison of P-wave velocity, S-wave velocity and density models respectively. The inverted P-wave velocity and S-wave velocity models generally match the well log data. However, it appears that the shallow parts of inverted density model (Figure 21f) by traditional simultaneous inversion strategy are underestimated, as indicated by the shaded area in Figure 22c. Furthermore, artifacts appear in the deep parts of the inverted density model as indicated by the arrows in Figures 21f and 22c. In the inverted density model (21i) by new inversion method, the shallow parts are better recovered and the artifacts in deep parts are suppressed. The new inversion method also provides faster convergence as shown in Figure 23.

For resolution analysis, in Figure 24a, approximate eigenvalue volumes $\mathbf{Eig}_{\alpha\alpha}$, $\mathbf{Eig}_{\beta\beta}$ and $\mathbf{Eig}_{\rho'\rho'}$ of the inverted P-wave velocity, S-wave velocity and density models by traditional simultaneous inversion method are plotted. In Figure 24b, the approximate eigenvalue volumes of the inverted models generated using the new inversion method are given.

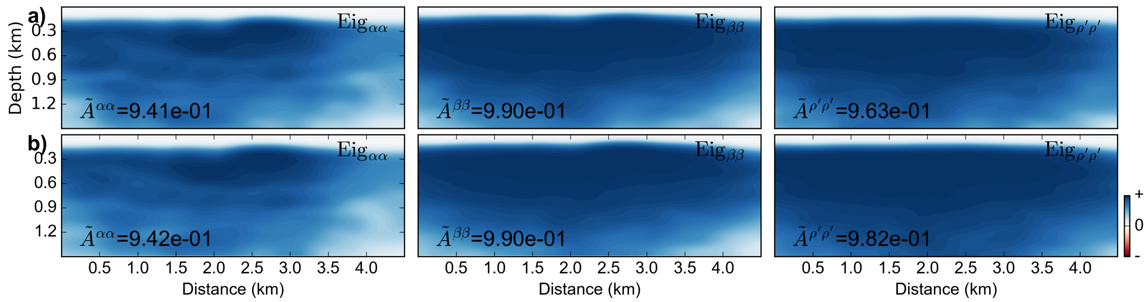

FIG. 24. (a) illustrate the approximate eigenvalue volumes $\mathbf{Eig}_{\alpha\alpha}$, $\mathbf{Eig}_{\beta\beta}$ and $\mathbf{Eig}_{\rho'\rho'}$ of the inverted models by traditional simultaneous inversion method; (b) illustrate the approximate eigenvalue volumes $\mathbf{Eig}_{\alpha\alpha}$, $\mathbf{Eig}_{\beta\beta}$ and $\mathbf{Eig}_{\rho'\rho'}$ of the inverted models generated using the new inversion method.

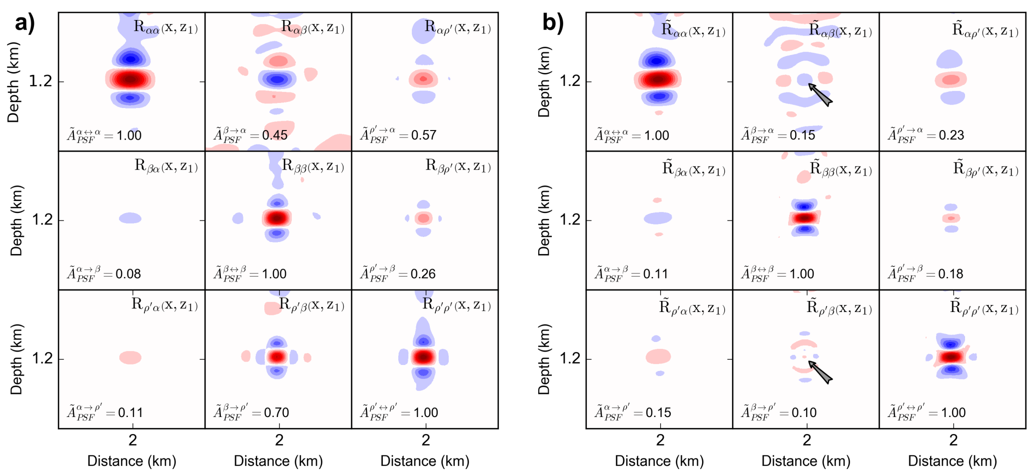

FIG. 25. (a) show the collection of EMPSFs at local position $\mathbf{z}_1 = (2.0\text{ km}, 1.2\text{ km})$ with conjugate-gradient algorithm; (b) show the collection of EMPSFs at $\mathbf{z}_1$ with conjugate-gradient algorithm following the work-flow of new inversion method (Table 4.1).

Both of these methods are able to recover amplitudes of the model parameters very well. The spatial and interparameter tradeoffs of the inverted models at location of $\mathbf{z}_1$=$(2.0\text{ km}, 1.2\text{ km})$ are evaluated, as indicated by blue square in Figure 21a. Figure 25a show the normalized EMPSFs with 10 conjugate-gradient iterations. Figure 25b show the normalized EMPSFs with 10 conjugate-gradient iterations following the work-flow of new inversion method. In Figure 25a, it is observed that the EMPSFs $\mathbf{R}_{\alpha\beta}(\mathbf{x}, \mathbf{z})$ and $\mathbf{R}_{\rho'\beta}(\mathbf{x}, \mathbf{z})$ representing the contaminations from S-wave velocity to P-wave velocity and density are still strong, which means the inverted P-wave velocity and density models (Figures 21d and 21f) by traditional simultaneous inversion method still suffer interparameter contaminations. In Figure 25b, the EMPSFs $\mathbf{R}_{\alpha\beta}(\mathbf{x}, \mathbf{z})$ and $\mathbf{R}_{\rho'\beta}(\mathbf{x}, \mathbf{z})$ are very weak, which means that with the new inversion method, the interparameter contaminations from S-wave velocity to P-wave velocity and density have been reduced. These numerical experiments show that the new inversion method is able to provide high resolution P-wave velocity and S-wave velocity models. Furthermore, the interparameter contaminations from S-wave velocity to P-wave velocity and density can be suppressed, which will provide more convincing isotropic-elastic parameters for reservoir characterization.

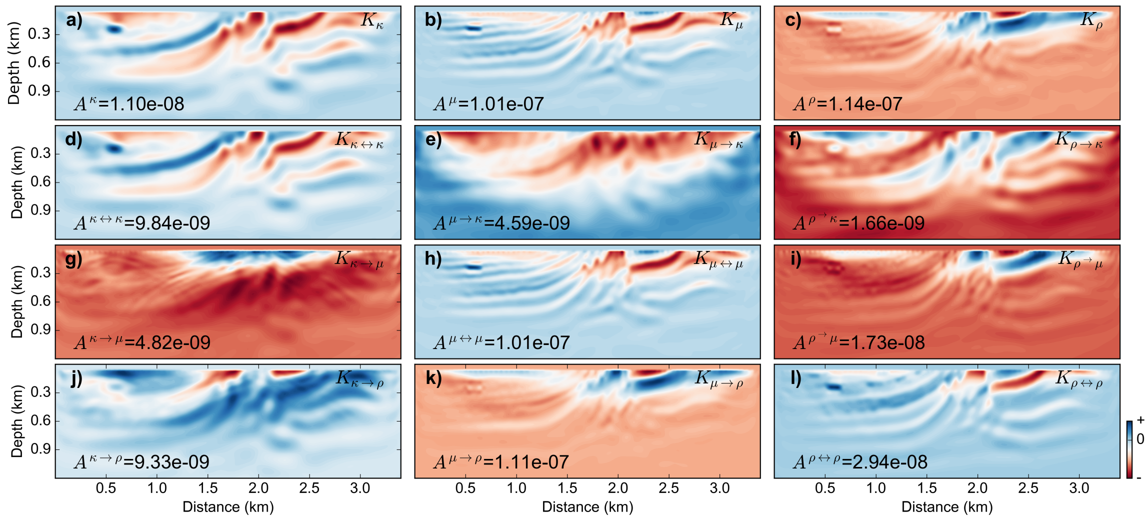

FIG. 26. (a-c) illustrate the standard sensitivity kernels K_κ, K_μ, and K_ρ; (d-f) show the correct update kernel $K_{\kappa\leftrightarrow\kappa}$ and contamination kernels $K_{\mu\to\kappa}$ and $K_{\rho\to\kappa}$; (g-i) show the contamination kernel $K_{\kappa\to\mu}$, correct update kernels $K_{\mu\leftrightarrow\mu}$ and contamination kernel $K_{\rho\to\mu}$; (j-l) show contamination kernels $K_{\kappa\to\rho}$ and $K_{\mu\to\rho}$ and correct update kernel $K_{\rho\leftrightarrow\rho}$. A represent maximum magnitudes of the kernels.

The influence of different parameterizations in isotropic-elastic FWI

Inversion experiments with velocity-density, modulus-density and impedance-density parameterizations are then carried out for comparison. The true models and initial models are shown in Figure 7. Figure 26 show the standard sensitivity kernels and true *interparameter contamination kernels* among bulk modulus κ, shear modulus μ and density ρ within modulus-density parameterization. It can be seen that the standard sensitivity kernel K_ρ is quite different from the correct update kernel $K_{\rho\leftrightarrow\rho}$. However, the contamination kernel $K_{\mu\to\rho}$ is very similar to the standard sensitivity kernel K_ρ. This is strongly suggestive that the update for density ρ is dominated by the contaminations from shear modulus μ to density ρ. Figure 27 shows the standard sensitivity kernels and true *interparameter contamination kernels* among P-wave impedance IP, S-wave impedance IS and density ρ'' within impedance-density parameterization. Similarly, the standard update kernel for density ρ'' is dominated by the contamination from S-wave impedance to density ρ''.

The inversion experiments are carried out using three frequency bands of [3Hz, 5Hz], [3Hz, 8Hz] and [3Hz, 10Hz] with 60 iterations at each frequency band. The same true models and initial models are used for inversion with the 3 different parameterizations. Figures 28, 29 and 30 show the corresponding true models, initial models and inverted models after 180 iterations. Note: here I use non-linear conjugate-gradient method for inversion. In Figure 31, the convergence histories at frequency band of [3Hz, 5Hz] are plotted for comparison. In Figure 28, P-wave velocity and S-wave velocity are well reconstructed. The density structures are also reconstructed even though there are still some interparameter contaminations. In Figure 29, the shear modulus μ is recovered best. The inverted density model is distorted, which may be caused by the contaminations from shear modulus. Magnitudes of the recovered bulk modulus model are very weak. In Figure 30, the reconstructed density structures are distorted significantly, which may be caused by

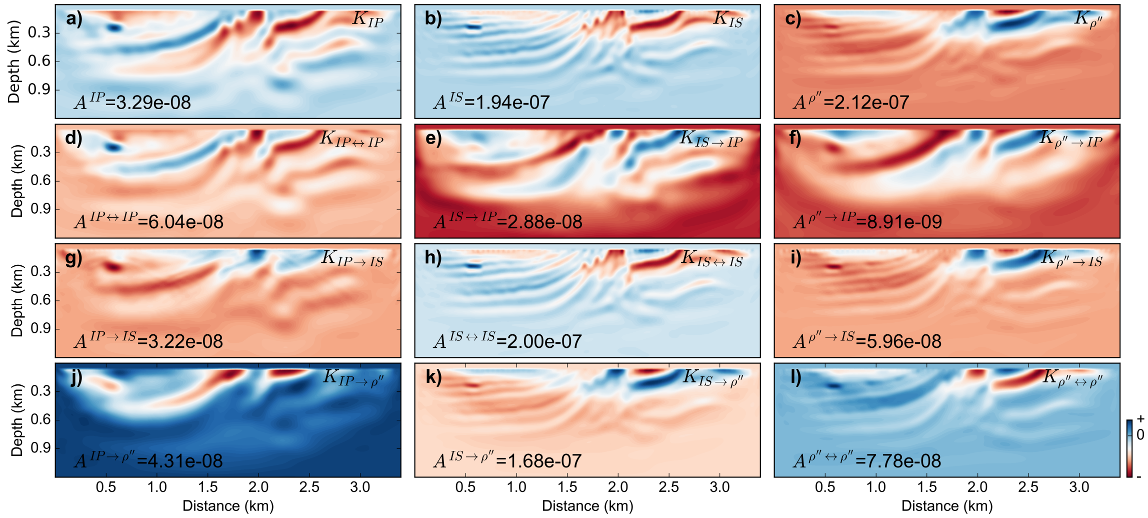

FIG. 27. (a-c) illustrate the standard sensitivity kernels K_{IP}, K_{IS}, and $K_{\rho''}$; (d-f) show the correct update kernel $K_{\text{IP}\leftrightarrow\text{IP}}$ and contamination kernels $K_{\text{IS}\rightarrow\text{IP}}$ and $K_{\rho''\rightarrow\text{IP}}$; (g-i) show the contamination kernel $K_{\text{IP}\rightarrow\text{IS}}$, correct update kernels $K_{\text{IS}\leftrightarrow\text{IS}}$ and contamination kernel $K_{\rho''\rightarrow\text{IS}}$; (j-l) show contamination kernels $K_{\text{IP}\rightarrow\rho''}$ and $K_{\text{IS}\rightarrow\rho''}$ and correct update kernel $K_{\rho''\leftrightarrow\rho''}$. A represent maximum magnitudes of the kernels.

the strong interparameter contaminations from S-wave impedance. This observation also verifies our analysis with scattering patterns in Figure **??**. Because the scattered wavefields due to density perturbations in impedance-density parameterization mostly forward scattered, recorded data on surface is mainly caused by P-wave impedance and S-wave impedance perturbations, which makes it more difficult to recover density structures. In Figure 31, velocity-density parameterization provides the fastest convergence rate. Hence, velocity-density parameterization is still the best choice to recover isotropic-elastic parameters among these three parameterizations.

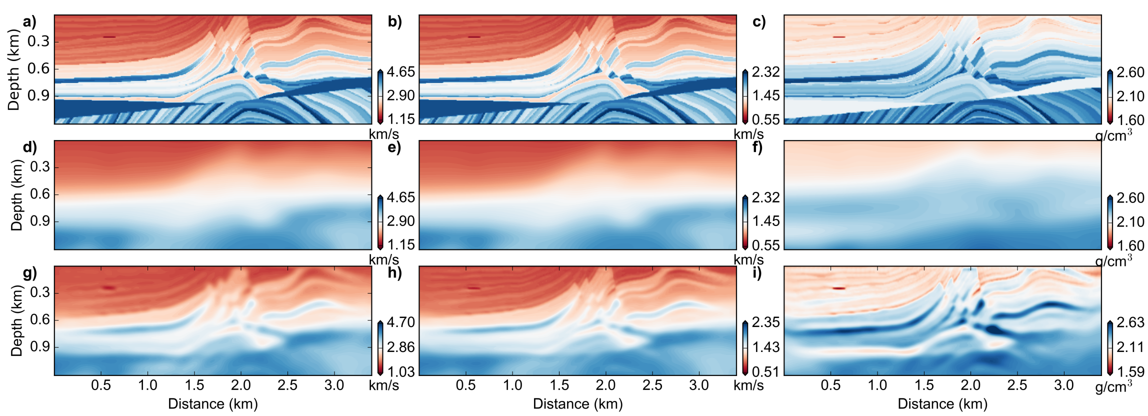

FIG. 28. (a-c) show the true P-wave velocity, S-wave velocity and density models; (d-f) show the corresponding initial P-wave velocity, S-wave velocity and density models; (g-i) show the corresponding inverted P-wave velocity, S-wave velocity and density models.

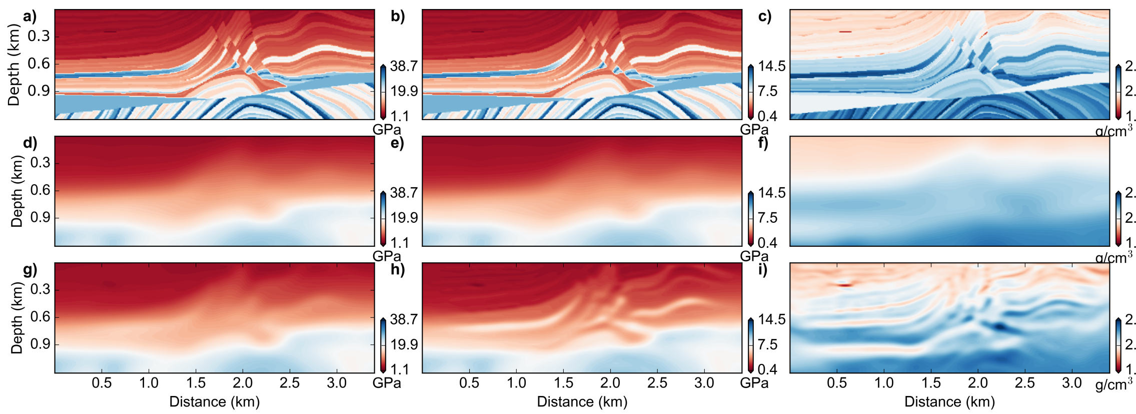

FIG. 29. (a-c) show the true bulk modulus κ, shear modulus μ and density ρ models; (d-f) show the corresponding initial bulk modulus κ, shear modulus μ and density ρ models; (g-i) show the corresponding inverted bulk modulus κ, shear modulus μ and density ρ models.

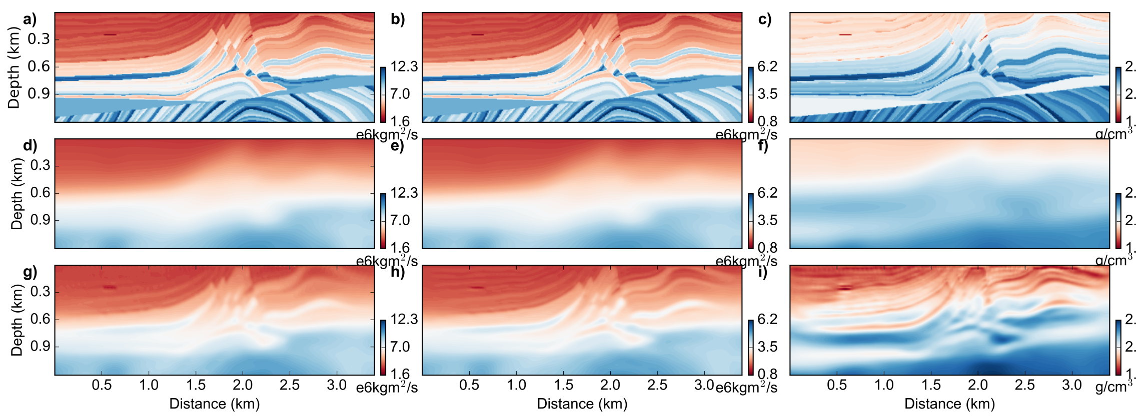

FIG. 30. (a-c) show the true P-wave impedance IP, S-wave impedance IS and density ρ'' models; (d-f) show the corresponding initial P-wave impedance IP, S-wave impedance IS and density ρ'' models; (g-i) show the corresponding inverted P-wave impedance IP, S-wave impedance IS and density ρ'' models.

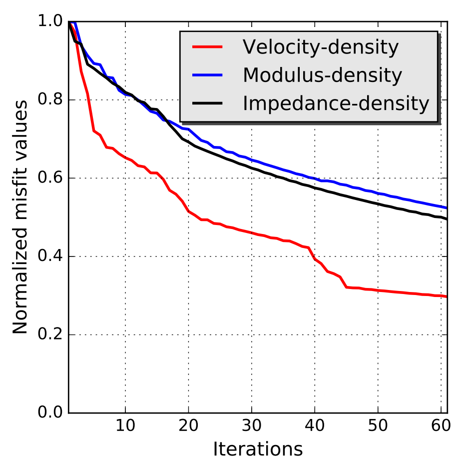

FIG. 31. Convergence rates comparison for various parameterizations in isotropic-elastic FWI ([3Hz, 5Hz] frequency band). The red, blue and black curves indicate velocity-density, modulus-density and impedance-density parameterization respectively.

DISCUSSION

Interparameter tradeoffs are strongly influenced by source-receive illumination (or acquisition geometry). In this chapter, the interparameter tradeoffs with perfect acquisition geometry are studied using a simple Gaussian-anomaly model and reflection acquisition geometry using Marmousi model. In transmission tomography (i.e., cross-well survey or earthquake seismology), the strengths and characteristics of the interparameter contaminations may be different from the conclusions and results presented in this chapter. Hence, for inverse problems with different models and acquisition geometries, the interparameter tradeoffs should be reevaluated following the strategies presented in this chapter.

Various misfit functions (i.e., envelope, instantaneous phase and traveltime misfit functions) based on different measurements have been studied for full-waveform inversion (Bozdag et al., 2011). Different physical parameters are sensitive to different measurements (i.e., amplitude and traveltime). The interparameter tradeoffs with the common waveform difference based misfit function are studied. It is also necessary to assess the interparameter tradeoffs in isotropic-elastic FWI for different misfit functions.

CONCLUSIONS

Origins of interparameter tradeoffs in isotropic-elastic FWI have been revealed with *interparameter contamination kernels*. Strengths and characteristics of the interparameter contaminations in isotropic-elastic FWI are quantified locally or within the whole volume by applying multiparameter Hessian to various types of probes. Two approaches (adjoint-state and finite-difference) are examined to construct the multiparameter Hessian matrix-vector products. This chapter reveals that S-wave velocity perturbations produce relatively strong contaminations into density updates and phase-revered contaminations into P-wave velocity updates. These contaminations make density structures highly under- and overestimated.

A novel inversion strategy has been recommended to reduce the contaminations from S-wave velocity to other parameters based on approximate contamination kernels. Numerical examples are given to illustrate that this new inversion strategy is able to provide more convincing and reliable density estimations in isotropic-elastic FWI. Approximate eigenvalue volume is employed to evaluate resolution of inverted models within the whole volume. Both of local spatial and interparameter tradeoffs of the inverted models are evaluated with extended multiparameter point spread functions (EMPSFs), which provide more accurate measurements of the local resolution compared to traditional MPSFs. The proposed strategies are finally applied on Hussar practical seismic dataset. According to the inverted models with different parameterizations, it is concluded that the velocity-density parameterization is still a better choice than modulus-density and impedance-density parameterizations.

ACKNOWLEDGEMENTS

This research was supported by the Consortium for Research in Elastic Wave Exploration Seismology (CREWES) and National Science and Engineering Research Council of Canada (NSERC, CRDPJ 461179-13). Wenyong Pan is also supported by SEG/Chevron

Scholarship and Eyes High International Doctoral Scholarship. Thanks also to Tiger cluster (Princeton University) and Lattice cluster (Compute Canada) for providing parallel computing facilities. Thanks specially to Frederik Simons and Yanhua O. Yuan for constructive comments and suggestions when I visited Princeton University in 2016. Thanks also to Andreas Fichtner, Samual Gray, Daniel Trad, Junxiao Li, Hassan Khaniani, Youyi Ruan, Wenjie Lei, Ryan Modrak and James Smith for valuable discussions.

APPENDIX A: EXPLICIT EXPRESSIONS OF INTERPARAMETER CONTAMINATION KERNELS IN ISOTROPIC-ELASTIC FWI

In this appendix, explicit expressions of correct update kernels and *interparameter contamination kernels* in equations (29) and (30) are given. $K_{\beta\leftrightarrow\beta}$ and $K_{\rho'\leftrightarrow\rho'}$ are correct update kernels for β and ρ':

$$\begin{aligned} K_{\beta\leftrightarrow\beta} &= -a_\beta^{-1}\int_{\Omega(\mathbf{x}')} H_{\beta\beta}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_\beta\left(\mathbf{x}'\right)d\mathbf{x}' \\ &= -\langle 2\rho'\beta^2\left[\partial_j G_{ni}\left(\mathbf{x}\right)\left(\partial_i u_j\left(\mathbf{x}\right)+\partial_j u_i\left(\mathbf{x}\right)\right)-2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right] \\ &\quad\times 2\rho'\beta^2\left[\partial_{j'} G_{n'i'}\left(\mathbf{x}'\right)\Delta m_\beta\left(\mathbf{x}'\right)\left(\partial_{i'} u_{j'}\left(\mathbf{x}'\right)+\partial_{j'} u_{i'}\left(\mathbf{x}'\right)\right)\right. \\ &\quad\left.-2\partial_{i'} G_{n'i'}\left(\mathbf{x}'\right)\Delta m_\beta\left(\mathbf{x}'\right)\partial_{k'} u_{k'}\left(\mathbf{x}'\right)\right]\rangle, \end{aligned} \tag{66}$$

$$\begin{aligned} K_{\rho'\leftrightarrow\rho'}\left(\mathbf{x}\right) &= -a_{\rho'}^{-1}\int_{\Omega(\mathbf{x}')} H_{\rho'\rho'}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)d\mathbf{x}' \\ &= -\langle \rho'\left[\left(G_{ni}\left(\mathbf{x}\right)\partial_t^2 u_i\left(\mathbf{x}\right)+2\alpha^2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right)\right. \\ &\quad\left.+2\beta^2\left(\partial_j G_{ni}\left(\mathbf{x}\right)\left(\partial_i u_j\left(\mathbf{x}\right)+\partial_j u_i\left(\mathbf{x}\right)\right)-2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right)\right] \\ &\quad\times\rho'\left[\left(G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{t'}^2 u_{i'}\left(\mathbf{x}'\right)+2\alpha^2\partial_{i'} G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'} u_{k'}\left(\mathbf{x}'\right)\right)\right. \\ &\quad+2\beta^2\left(\partial_{j'} G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\left(\partial_{i'} u_{j'}\left(\mathbf{x}'\right)+\partial_{j'} u_{i'}\left(\mathbf{x}'\right)\right)\right. \\ &\quad\left.\left.-2\partial_{i'} G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'} u_{k'}\left(\mathbf{x}'\right)\right)\right]\rangle. \end{aligned} \tag{67}$$

Interparameter contamination kernels $K_{\alpha\to\beta}$ and $K_{\rho'\to\beta}$ represent the mappings from α and ρ' to β respectively:

$$\begin{aligned} K_{\alpha\to\beta} &= -a_\beta^{-1}\int_{\Omega(\mathbf{x}')} H_{\beta\alpha}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_\alpha\left(\mathbf{x}'\right)d\mathbf{x}' \\ &= -\langle 2\rho'\beta^2\left[\partial_j G_{ni}\left(\mathbf{x}\right)\left(\partial_i u_j\left(\mathbf{x}\right)+\partial_j u_i\left(\mathbf{x}\right)\right)-2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right] \\ &\quad\times\left[2\rho'\alpha^2\partial_{i'} G_{i'n'}\left(\mathbf{x}'\right)\Delta m_\alpha\left(\mathbf{x}'\right)\partial_{k'} u_{k'}\left(\mathbf{x}'\right)\right]\rangle, \end{aligned} \tag{68}$$

$$\begin{aligned} K_{\rho'\to\beta} &= -a_\beta^{-1}\int_{\Omega(\mathbf{x}')} H_{\beta\rho'}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)d\mathbf{x}' \\ &= -\langle 2\rho'\beta^2\left[\partial_j G_{ni}\left(\mathbf{x}\right)\left(\partial_i u_j\left(\mathbf{x}\right)+\partial_j u_i\left(\mathbf{x}\right)\right)-2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right] \\ &\quad\times\rho'\left[\left(G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{t'}^2 u_{i'}\left(\mathbf{x}'\right)+2\alpha^2\partial_{i'} G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'} u_{k'}\left(\mathbf{x}'\right)\right)\right. \\ &\quad+2\beta^2\left(\partial_{j'} G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\left(\partial_{i'} u_{j'}\left(\mathbf{x}'\right)+\partial_{j'} u_{i'}\left(\mathbf{x}'\right)\right)\right. \\ &\quad\left.\left.-2\partial_{i'} G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\rho'}\left(\mathbf{x}'\right)\partial_{k'} u_{k'}\left(\mathbf{x}'\right)\right)\right]\rangle. \end{aligned} \tag{69}$$

Interparameter contamination kernels $K_{\alpha\to\rho'}$ and $K_{\beta\to\rho'}$ represent the mappings from α and β to ρ' respectively:

$$\begin{aligned} K_{\alpha\to\rho'} &= -a_{\rho'}^{-1}\int_{\Omega(\mathbf{x}')} H_{\rho'\alpha}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_{\alpha}\left(\mathbf{x}'\right)d\mathbf{x}' \\ &= -\langle\rho'\left[\left(G_{ni}\left(\mathbf{x}\right)\partial_t^2 u_i\left(\mathbf{x}\right)+2\alpha^2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right)\right. \\ &\left.+2\beta^2\left(\partial_j G_{ni}\left(\mathbf{x}\right)\left(\partial_i u_j\left(\mathbf{x}\right)+\partial_j u_i\left(\mathbf{x}\right)\right)-2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right)\right] \\ &\times\left[2\rho'\alpha^2\partial_{i'}G_{i'n'}\left(\mathbf{x}'\right)\Delta m_{\alpha}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right]\rangle, \end{aligned} \tag{70}$$

$$\begin{aligned} K_{\beta\to\rho'} &= -a_{\rho'}^{-1}\int_{\Omega(\mathbf{x}')} H_{\rho'\beta}\left(\mathbf{x},\mathbf{x}'\right)\Delta m_{\beta}\left(\mathbf{x}'\right)d\mathbf{x}' \\ &= -\langle\rho'\left[\left(G_{ni}\left(\mathbf{x}\right)\partial_t^2 u_i\left(\mathbf{x}\right)+2\alpha^2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right)\right. \\ &\left.+2\beta^2\left(\partial_j G_{ni}\left(\mathbf{x}\right)\left(\partial_i u_j\left(\mathbf{x}\right)+\partial_j u_i\left(\mathbf{x}\right)\right)-2\partial_i G_{ni}\left(\mathbf{x}\right)\partial_k u_k\left(\mathbf{x}\right)\right)\right] \\ &\times\rho'\beta^2\left[\partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\beta}\left(\mathbf{x}'\right)\left(\partial_{i'}u_{j'}\left(\mathbf{x}'\right)+\partial_{j'}u_{i'}\left(\mathbf{x}'\right)\right)\right. \\ &\left.-2\partial_{i'}G_{n'i'}\left(\mathbf{x}'\right)\Delta m_{\beta}\left(\mathbf{x}'\right)\partial_{k'}u_{k'}\left(\mathbf{x}'\right)\right]\rangle. \end{aligned} \tag{71}$$

APPENDIX B:MULTIPARAMETER HESSIAN-VECTOR PRODUCT CALCULATION IN TIME DOMAIN

Constructing Hessian-vector products is an essential step for implementing Hessian-free optimization methods and quantifying uncertainties of the inverse problems (Santosa and Symes, 1988; Fichtner and Trampert, 2011a; Métivier et al., 2012; Pan et al., 2017a). One popular approach for Hessian-vector product construction in time domain is known as adjoint-state method. To understand the mechanism of the adjoint-state method for Hessian-vector calculation, I first consider minimizing the misfit function (Métivier et al., 2013):

$$\Psi\left(\mathbf{m}\right)=\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T u_n^*\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)\boldsymbol{\nu}dt', \tag{72}$$

where $\boldsymbol{\nu}$ is an arbitrary function and the gradient is given by $\nabla_{\mathbf{m}}\Psi=\nabla_{\mathbf{m}}\mathbf{u}^*\boldsymbol{\nu}$. Minimizing this misfit function subject to that wavefield $\mathbf{u}$ satisfies the wave equation gives the augmented Lagrangian functional (Liu et al., 2006; Métivier et al., 2013):

$$\begin{aligned} \boldsymbol{\chi}\left(\mathbf{m},\mathbf{u},\boldsymbol{\lambda}\right)=&\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T\left[u_n^*\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)\boldsymbol{\nu}\right. \\ &\left.-\lambda_i\left(\rho\partial_{t'}^2 u_i\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)-\partial_j\left(c_{ijkl}\partial_l u_k\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)\right)-f_i\right)\right]dt', \end{aligned} \tag{73}$$

where f_i is the source term and λ_i is the Lagrangian multiplier. Variation of functional due to the perturbations of model parameter $\Delta\mathbf{m}$ and wavefield $\Delta\mathbf{u}$ is given by (Liu et al., 2006):

$$\begin{aligned} \Delta\boldsymbol{\chi}\left(\mathbf{m},\mathbf{u},\boldsymbol{\lambda}\right)=&\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_0^T-\left[\Delta\rho\lambda_i\partial_{t'}^2 u_i\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)+\Delta c_{ijkl}\partial_j\lambda_i\left(\partial_l u_k\left(\mathbf{x}_g,\mathbf{x}_s,t'\right)\right)\right] \\ &+\left[\boldsymbol{\nu}-\left(\rho\partial_{t'}^2\lambda_i-\partial_j\left(c_{ijkl}\partial_l\lambda_k\right)\right)\right]\Delta u_n dt'. \end{aligned} \tag{74}$$

Equation (74) is stationary with respect to wavefield perturbation $\Delta\mathbf{u}$ when its coefficient is zero, which gives the adjoint-state equation: $\rho\partial_{t'}^2\lambda_i - \partial_j\left(c_{ijkl}\partial_l\lambda_k\right) = \boldsymbol{\nu}$, where $\boldsymbol{\nu}$ serves as the adjoint source. Thus, gradients of the misfit function with respect to density ρ and elastic constants $\mathbf{c}$ become:

$$\nabla_\rho\boldsymbol{\chi} = \nabla_\rho\mathbf{u}^*\boldsymbol{\nu} = -\langle\tilde{u}_i\partial_{t'}^2 u_i\rangle, \nabla_\mathbf{c}\boldsymbol{\chi} = \nabla_\mathbf{c}\mathbf{u}^*\boldsymbol{\nu} = -\langle\partial_j\tilde{u}_i\partial_l u_k\rangle, \tag{75}$$

where $\tilde{u}_i\left(\mathbf{x}_g, \mathbf{x}, T-t'\right) = \lambda_i\left(\mathbf{x}_g, \mathbf{x}, T-t'\right) = G_{ni}\left(\mathbf{x}_g, \mathbf{x}, T-t'\right)\boldsymbol{\nu}$ is the adjoint wavefield. Product of Jacobian matrix with an arbitrary vector $\mathbf{v}$ is given by:

$$\mathfrak{J} = \nabla_\mathbf{m}\mathbf{u}\mathbf{v} = -\langle G_{n'i'}\left(\mathbf{x}'\right)v\left(\mathbf{x}'\right)\partial_{t''}^2 u_{i'}\left(\mathbf{x}'\right) + \partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)v\left(\mathbf{x}'\right)\partial_{l'}u_{k'}\left(\mathbf{x}'\right)\rangle, \tag{76}$$

Replacing $\boldsymbol{\nu}$ in equation (75) with Jacobian-vector product $\mathfrak{J}$ gives the multiparameter Gauss-Newton Hessian-vector product:

$$\begin{aligned}\mathfrak{H} &= \nabla_\mathbf{m}\mathbf{u}^*\nabla_\mathbf{m}\mathbf{u}\mathbf{v}\\ &= \langle\left(G_{ni}\left(\mathbf{x}\right)\partial_{t'}^2 u_i\left(\mathbf{x}\right) + \partial_j G_{ni}\left(\mathbf{x}\right)\partial_l u_k\left(\mathbf{x}\right)\right)\\ &\times\left(G_{n'i'}\left(\mathbf{x}'\right)v\left(\mathbf{x}'\right)\partial_{t''}^2 u_{i'}\left(\mathbf{x}'\right) + \partial_{j'}G_{n'i'}\left(\mathbf{x}'\right)v\left(\mathbf{x}'\right)\partial_{l'}u_{k'}\left(\mathbf{x}'\right)\right)\rangle.\end{aligned} \tag{77}$$

For example, product of off-diagonal block $\mathbf{H}_{\mathbf{c}\rho}$ in multiparameter Gauss-Newton Hessian with an arbitrary perturbation vector $\mathbf{v}_\rho$ due to density can be written explicitly as:

$$\begin{aligned}\mathfrak{H}_{\mathbf{c}\rho}\left(\mathbf{x}\right) &= \int_{\Omega(\mathbf{x}')} H_{\mathbf{c}\rho}\left(\mathbf{x}, \mathbf{x}'\right)v_\rho\left(\mathbf{x}'\right)d\mathbf{x}'\\ &= \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_{\Omega(\mathbf{x}')}\int\int\partial_k u_l\left(\mathbf{x}, \mathbf{x}_s, t'\right)\partial_j G_{ni}\left(\mathbf{x}_g, \mathbf{x}, T-t'\right)\\ &\times G_{n'i'}\left(\mathbf{x}_g, \mathbf{x}', t-t''\right)v_\rho\left(\mathbf{x}'\right)\partial_{t''}^2 u_{i'}\left(\mathbf{x}', \mathbf{x}_s, t''\right)dt'dt''d\mathbf{x}',\end{aligned} \tag{78}$$

where Jacobian-vector product $\mathfrak{J}_\rho$ is:

$$\mathfrak{J}_\rho\left(\mathbf{x}_g, \mathbf{x}', t\right) = \sum_{\mathbf{x}_s}\int G_{n'i'}\left(\mathbf{x}_g, \mathbf{x}', t-t''\right)v_\rho\left(\mathbf{x}'\right)\partial_{t''}^2 u_{i'}\left(\mathbf{x}', \mathbf{x}_s, t''\right)dt'', \tag{79}$$

where interaction of the indicate wavefield $u_{i'}\left(\mathbf{x}', \mathbf{x}_s, t''\right)$ with the perturbation vector $v_\rho\left(\mathbf{x}'\right)$ serves as the "secondary scattered source" $f_i'\left(\mathbf{x}', \mathbf{x}_s, t''\right)$:

$$f_i'\left(\mathbf{x}', \mathbf{x}_s, t''\right) = v_\rho\left(\mathbf{x}'\right)\partial_{t''}^2 u_{i'}\left(\mathbf{x}', \mathbf{x}_s, t''\right). \tag{80}$$

Convolution of this scattered source with the Green's function $G_{n'i'}\left(\mathbf{x}_g, \mathbf{x}', t-t''\right)$ gives the first-order scattered wavefield:

$$\Delta u_{n'}\left(\mathbf{x}_g, \mathbf{x}', t\right) = \sum_{\mathbf{x}_s}\int G_{n'i'}\left(\mathbf{x}_g, \mathbf{x}', t-t''\right)f_i'\left(\mathbf{x}', \mathbf{x}_s, t''\right)dt''. \tag{81}$$

Recorded scattered wavefield at the receiver locations can be considered as adjoint source $\tilde{f}_n'\left(\mathbf{x}, \mathbf{x}', t\right)$:

$$\tilde{f}_n'\left(\mathbf{x}, \mathbf{x}', t\right) = \sum_{\mathbf{x}_g}\Delta u_n\left(\mathbf{x}_g, \mathbf{x}', T-t\right)\delta\left(\mathbf{x}-\mathbf{x}_g\right). \tag{82}$$

Inserting equation (82) into equation (78) gives the Hessian-vector product as:

$$\mathfrak{H}_{\mathbf{c}\rho}\left(\mathbf{x}\right)=\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\int_{\Omega(\mathbf{x}')}\int\partial_k u_l\left(\mathbf{x},\mathbf{x}_s,t'\right)\partial_j\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},\mathbf{x}',T-t'\right)dt'd\mathbf{x}', \tag{83}$$

where $\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},\mathbf{x}',T-t'\right)$ is the adjoint wavefield:

$$\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},\mathbf{x}',T-t'\right)=\int_0^{T-t'}G_{ni}\left(\mathbf{x}_g,\mathbf{x},T-t'-t'''\right)\tilde{f}'_n\left(\mathbf{x},\mathbf{x}',t'''\right)dt'''. \tag{84}$$

Calculating Gauss-Newton Hessian-vector product with the adjoint-state approach needs to construct forward wavefield $u_{i'}\left(\mathbf{x}',\mathbf{x}_s,t''\right)$, Born modelling wavefield $\Delta u_{n'}\left(\mathbf{x}_g,\mathbf{x}',t\right)$ and adjoint wavefield $\tilde{u}_i\left(\mathbf{x}_g,\mathbf{x},\mathbf{x}',T-t'\right)$, The computational cost is 1.5 times than that of calculating gradient (Métivier et al., 2013).

Another approach for Hessian-vector calculation is finite-difference method. Recalling that Hessian operator represents the Fréchet derivative of the gradient vector, with Taylor series expansion:

$$\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}^0+\Delta\mathbf{m}\right)\approx\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}^0\right)+\mathbf{H}\Delta\mathbf{m}, \tag{85}$$

where $\mathbf{m}^0$ denote current model. Replacing the model perturbation vector $\Delta\mathbf{m}$ with an arbitrary vector $\mathbf{v}$ scaled by a small constant value $\bar{\epsilon}$ gives:

$$\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}^0+\bar{\epsilon}\mathbf{v}\right)\approx\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}^0\right)+\bar{\epsilon}\mathbf{H}\mathbf{v}. \tag{86}$$

An approximate Hessian-vector product solution can be obtained by:

$$\mathfrak{H}\approx\frac{\nabla_{\mathbf{m}}\Phi\left(\mathbf{m}^0+\bar{\epsilon}\mathbf{v}\right)-\nabla_{\mathbf{m}^0}\Phi\left(\mathbf{m}\right)}{\bar{\epsilon}}. \tag{87}$$

Two additional pairs of forward and adjoint simulations are required for calculating this Hessian-vector product approximation, which is affordable for large-scale inverse problems. Even this approximation may suffer from rounding errors, the accuracy can be improved with high-order finite-difference approaches at the cost of more computation requirements and for very small coefficient $\bar{\epsilon}$, its accuracy will be very high. For example, if the multiparameter Hessian is applied to vector $\mathbf{v}=[\mathbf{v}_{\mathbf{c}}=0\ \ \mathbf{v}_\rho\neq 0]^\dagger$, the Hessian-vector product $\mathfrak{H}_{\mathbf{c}\rho}=\mathbf{H}_{\mathbf{c}\rho}\mathbf{v}_\rho$ can be obtained by:

$$\mathfrak{H}_{\mathbf{c}\rho}\approx\frac{\nabla_{\mathbf{c}}\Phi\left(\rho_0+\bar{\epsilon}\mathbf{v}_\rho\right)-\nabla_{\mathbf{c}}\Phi\left(\rho_0\right)}{\bar{\epsilon}}. \tag{88}$$

Next, the two approaches with a multiparameter acoustic example are examined for comparison. Figure 32a shows the vector $\mathbf{v}_\rho$ with 9 isolated spikes. A homogeneous model with bulk modulus $\kappa=13.5$ GPa and density $\rho=1500$ kg/m^3 is used as the background model. A set of sources and receivers are distributed regularly along the top boundary of the model. Multiparameter Hessian-vector product $\mathfrak{H}_{\kappa\rho}=\mathbf{H}_{\kappa\rho}\mathbf{v}_\rho$ is calculated with adjoint-state and finite-difference methods, as shown in Figures 32b and 32c respectively. The Hessian-vector products by these two methods match very well. In this chapter, I adopt the adjoint-state approach for calculating the multiparameter Hessian-vector products in isotropic-elastic media.

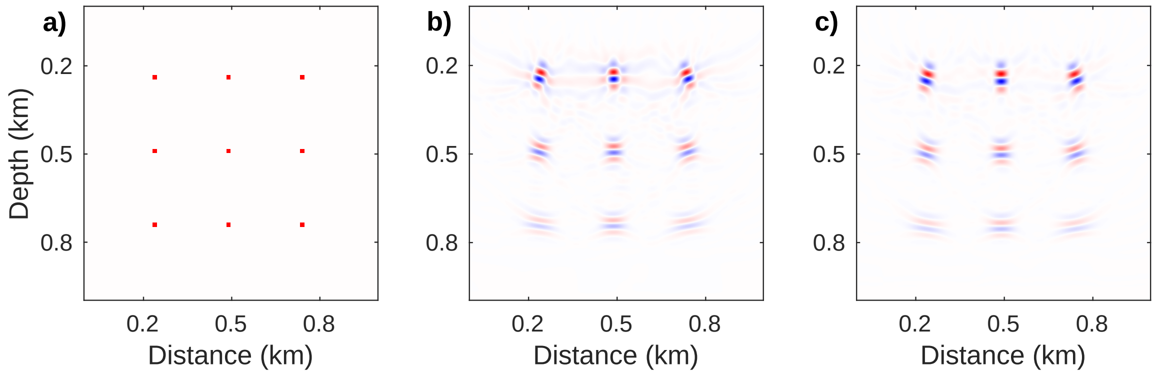

FIG. 32. Panel (a) shows vector $\mathbf{v}_\rho$ with 9 isolated spikes; Figures (b) and (c) are the multiparameter Hessian-vector products $\mathfrak{H}_{\kappa\rho} = \mathbf{H}_{\kappa\rho}\mathbf{v}_\rho$ calculated with second-order adjoint-state and finite-difference methods respectively.

APPENDIX C:STOCHASTICALLY ESTIMATING DIAGONALS OF MULTIPARAMETER HESSIAN OFF-DIAGONAL BLOCKS

This Appendix explains how to efficiently estimate the diagonals of multiparameter Hessian off-diagonal blocks with stochastic probing approach. Zero-mean random vector $\mathbf{v}$ is divided into N_p independent subvectors corresponding to N_p different physical parameters: $\mathbf{v} = \left[\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{N_p}\right]$. $\mathbf{v}_p \sim \mathcal{N}\left(\mathbb{E}\left[\mathbf{v}_p\right], \mathbf{\Sigma}_{\mathbf{v}_p\mathbf{v}_p}\right)$ and $\mathbf{v}_q \sim \mathcal{N}\left(\mathbb{E}\left[\mathbf{v}_q\right], \mathbf{\Sigma}_{\mathbf{v}_q\mathbf{v}_q}\right)$ are independent subvectors within $\mathbf{v}$. Expectation values and variance-covariance matrices of the subvectors satisfy:

$$\mathbb{E}\left[v_p\left(\mathbf{x}\right)\right] = 0, \mathbb{E}\left[v_q\left(\mathbf{x}\right)\right] = 0, \tag{89}$$

$$\Sigma_{\mathbf{v}_p\mathbf{v}_p}\left(v_p\left(\mathbf{x}\right), v_p\left(\mathbf{x}'\right)\right) = \mathbb{E}\left[v_p\left(\mathbf{x}\right) v_p\left(\mathbf{x}'\right)\right] = \delta\left(\mathbf{x} - \mathbf{x}'\right), \tag{90}$$

$$\Sigma_{\mathbf{v}_q\mathbf{v}_q}\left(v_q\left(\mathbf{x}\right), v_q\left(\mathbf{x}'\right)\right) = \mathbb{E}\left[v_q\left(\mathbf{x}\right) v_q\left(\mathbf{x}'\right)\right] = \delta\left(\mathbf{x} - \mathbf{x}'\right). \tag{91}$$

Cross-covariance between $\mathbf{v}_p$ and $\mathbf{v}_p$ can be obtained as:

$$\begin{aligned}\Sigma_{\mathbf{v}_p\mathbf{v}_q}\left(v_p\left(\mathbf{x}\right), v_q\left(\mathbf{x}'\right)\right) &= \mathbb{E}\left[\left(v_p\left(\mathbf{x}\right) - \mathbb{E}\left[v_p\left(\mathbf{x}\right)\right]\right)\left(v_q\left(\mathbf{x}'\right) - \mathbb{E}\left[v_q\left(\mathbf{x}'\right)\right]\right)^\dagger\right] \\ &= \mathbb{E}\left[v_p\left(\mathbf{x}\right) v_q\left(\mathbf{x}'\right)\right] - \mathbb{E}\left[v_p\left(\mathbf{x}\right)\right]\left(\mathbb{E}\left[v_q\left(\mathbf{x}'\right)\right]\right)^\dagger \\ &= 0.\end{aligned} \tag{92}$$

Thus, expectation value of the correlation result between subvector $\mathbf{v}_p$ and the sub-Hessian-vector product $\mathfrak{H}_p$ is given by:

$$\begin{aligned}\mathbb{E}\left[\mathbf{v}_p \odot \mathfrak{H}_p\right] &= \sum_{q=1}^{N_p} \mathbf{v}_p \odot \mathbf{H}_{pq}\mathbf{v}_q \\ &= \int_{\Omega(\mathbf{x}')} H_{pq}\left(\mathbf{x}, \mathbf{x}'\right) \mathbb{E}\left[v_p\left(\mathbf{x}\right) v_p\left(\mathbf{x}'\right)\right] d\mathbf{x}' \\ &= \mathbf{H}_{pq}\left(\mathbf{x}, \mathbf{x}\right).\end{aligned} \tag{93}$$

REFERENCES

Alkhalifa, T., and Plessix, R., 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytic parameter resolution study: Geophysics, **79**, R91–R101.

An, M., 2012, A simple method for determing the spatial resolution of a general inverse problem: Geophysical Journal International, **191**, 849–864.

Backus, G., and Gilbert, F., 1968, The resolving power of gross earth data: Geophysical Journal International, **16**, 169–205.

Baumstein, A., 2014, Extended subspace method for attenuation of crosstalk in multi-parameter Full Wavefield Inversion: SEG Expanded Abstracts, 1121–1125.

Bernauer, M., Fichtner, A., and Igel, H., 2014, Optimal observables for multiparameter seismic tomography: Geophysical Journal International, **198**, 1241–1254.

Boehm, C., and Ulbrich, M., 2014, A Semismooth Newton-CG Method for Constrained Parameter Identification in Seismic Tomography: SIAM J. Sci. Comput, **37**, S334–S364.

Borisov, D., and Singh, S. C., 2015, Three-dimensional elastic full waveform inversion in a marine environment using multicomponent ocean-bottom cables: synthetic study: Geophysical Journal International, **201**, 1215–1234.

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

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., and Pugmire, D., 2016, Global adjoint tomography: first-generation model: Geophysical Journal International, **207**, 1739–1766.

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.

Chen, B., and Xie, X., 2015, An efficient method for broadband seismic illumination and resolution analysises: SEG Expanded Abstracts, 4227–4231.

Cui, T., 2015, Improving seismic-to-well ties: M.Sc. thesis, University of Calgary.

Demanet, L., Létourneau, P. D., Noumal, N., Calandra, H., and Chiu, J. S., 2012, Matrix probing: a randomized preconditioner for the wave-equation Hessian: Appl. Comput. Harmon. Anal, **32**, 155–168.

Dettmer, J., Dosso, S. E., and Holland, C. W., 2007, Uncertainty estimation in seismo-acoustic reflection travel-time inversion: J. Acoust. Soc. Am., **122**, 161–176.

Esmaeili, S., 2016, Influence of low frequencies on seismic impedance inversion: M.Sc. thesis, University of Calgary.

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

Fichtner, A., and Trampert, J., 2011b, Resolution analysis in full waveform inversion: Geophysical Journal International, **187**, 1604–1642.

Fichtner, A., and van Leeuwen, T., 2015, Resolution analysis by random probing: Journal of Geophysical Research: Solid Earth, **120**, 5549–5573.

Flath, H. P., Wilcox, L. C., Akcelik, V., Hill, J., Waanders, B. V. B., and Ghattas, O., 2011, Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations: SIAM Journal on Scientific Computing, **33**, 407–432.

Gholami, Y., Brossier, R., Operto, S., Prieus, V., Ribodetti, A., and Virieux, J., 2013a, Which parameterization is suitable for acoustic vertical transverse isotropic full waveform inversion? part 2: Synthetic and real data case studies from valhall: Geophysics, **78**, R107–R124.

Gholami, Y., Brossier, R., Operto, S., Ribodetti, A., and Virieux, J., 2013b, Which parametrization for acoustic VTI full waveform inversion? - part 1: sensitivity and trade-off analysis: Geophysics, **78**, R81–R105.

Gouveia, W., and Scales, J., 1998, Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis: Journal of Geophysical Research, **103**, 2759–2779.

Halko, N., Martinsson, P., and Tropp, J., 2011, Finding structure with randomness: probabilistic algorithms for constructing approximation matrix decomposition: SIAM Review, **53**, 217–281.

Hu, J., Schuster, G. T., and A.Valasek, P., 2001, Poststack migration deconvolution: Geophysics, **66**, 939–952.

Hutchinson, M. F., 1990, A stochastic estimator of the trace of the influence matrix for Laplacian smoothing filters: Communications in Statistics -Simulation and Computation, **19**, 433–450.

Innanen, K. A., 2013, Coupling in amplitude variation with offset and the wiggins approximation: Geophysics, **78**, N21–N33.

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

Kamath, N., and Tsvankin, I., 2014, Sensitivity anaysis for elastic full-waveform inversion in VTI media: SEG Expanded Abstracts, 1162–1166.

Kennett, B. L. N., Sambridge, M. S., and Williamson, P. R., 1988, Subspace methods for large inverse problems with multiple parameter classes: Geophysical Journal Interational, **94**, 237–247.

Köhn, D., Nil, D. D., Kurzmann, A., Przebindowska, A., and Bohlen, T., 2012, On the influence of model parameterizations in elastic full waveform inversion tomography: Geophysical Journal International, **191**, 325–345.

Komatitsch, D., and Tromp, J., 2005, Introduction to the spectral-element method for 3-D seismic wave propagation: Geophysical Journal Interational, **139**, 806–822.

Liu, F., Hanson, D. W., Whitemore, N. D., Day, R. S., and Stolt, R. H., 2006, Toward a unified analysis for source plane-wave migration: Geophysics, **71**, S129–S139.

Liu, Y., Yang, J., Chi, B., and Dong, L., 2015, An improved scattering-integral approach for frequency-domain full waveform inversion: Geophysical Journal International, **202**, 1827–1842.

Lloyd, H., 2013, An investigation of the role of low frequencies in seismic impedance inversion: M.Sc. thesis, University of Calgary.

Luo, Y., 2012, Seismic imaging and inversion based on spectral-element and adjoint methods: Ph.D. thesis, Princeton University.

Luo, Y., Tromp, J., Denel, B., and Calandra, H., 2013, 3D coupled acoustic-elastic migration with topography and bathymetry based on spectral-element and adjoint methods: Geophysics, **78**, S193–S202.

MacCarthy, J. K., Borchers, B., and Aster, R. C., 2011, Efficient stochastic estimation of the model resolution matrix diagonal and generalized cross-validation for large geophysical inverse problems: Journal of Geophysical Research, **116**, B10,304.

Margrave, G. F., Mewhort, L., Phillips, T., Hall, M., Bertram, M. B., Lawton, D. C., Innanen, K., Hall, K. W., and Bertram, K., 2012, The hussar low-frequency experiment: CSEG Recorder, 25–39.

Métivier, L., Brossier, R., Operto, S., and Virieux, J., 2015, Acoustic multi-parameter fwi for the reconstruction of p-wave velocity, density and attenuation: preconditioned truncated newton approach: SEG Expanded Abstracts, 1198–1203.

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.

Modrak, R., and Tromp, J., 2016, Seismic waveform inversion best practices: regional, global, and exploration best cases: Geophysical Journal Interational, **206**, 1864–1889.

Modrak, R., Tromp, J., and Yuan, Y. O., 2016, On the choice of materials parameters for elastic waveform inversion: SEG Expanded Abstracts, 1115–1119.

Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: Geophysics, **52**, 1211–1228.

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

Oh, J. W., and Alkhalifah, T., 2016, The scattering potential of partial derivative wavefields in 3-D elastic orthorhombic media: an inversion prospective: Geophysical Journal International, **206**, 1740–1760.

Oldenborger, G. A., and Routh, P. S., 2009, The point-spread function measure of resolution for the 3-Delectrical resistivity experiment: Geophysical Journal International, **176**, 405–414.

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., and Innanen, K. A., 2016a, Elastic full-waveform inversion: density effects, cycle-skipping, and inter-parameter mapping: CREWES Annual Report, 1–16.

Pan, W., and Innanen, K. A., 2016b, Suppress Parameter Cross-talk for Elastic Full-waveform Inversion: Parameterization and Acquisition Geometry: GeoConvention, 1–5.

Pan, W., and Innanen, K. A., 2016c, A Summary of Several Challenges Facing Multi-Parameter Elastic Full Waveform Inversion: CSEG Recorder, 36–39.

Pan, W., Innanen, K. A., and Liao, W., 2017a, Accelerating Hessian-free Gauss-Newton full-waveform inversion via l-BFGS preconditioned conjugate-gradient algorithm: Geophysics, **82**, R49–R64.

Pan, W., Innanen, K. A., Margrave, G. F., Fehler, M. C., Fang, X., and Li, J., 2016, 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., and Yuan, Y. O., 2017b, Inter-parameter tradeoff quantification and reduction in isotropic-elastic full-waveform inversion: SEG Expanded Abstracts, 1–5.

Pan, W., Innanen, K. A., and Yuan, Y. O., 2017c, Interparameter tradeoff quantification and reduction in isotropic-elastic full-waveform inversion: SEG Expanded Abstracts, 1539–1544.

Pan, W., Innanen, K. A., Yuan, Y. O., and Simons, F., 2017d, Quantifying parameter trade-off in elastic full-waveform inversion via multi-parameter Hessian probing: GeoConvention, 1–5.

Podgornova, O., Leaney, S., and Liang, L., 2015, Analysis of resolution limits of VTI anisotropy with full waveform inversion: SEG Expanded Abstracts, 1188–1192.

Raknes, E. B., and Arntsen, B., 2015, A numerical study of 3D elastic time-lapse full-waveform inversion using multicomponent seismic data: Geophysics, **80**, R303–R315.

Rawlinson, N., Fichtner, A., Sambridge, M., and K.Young, M., 2014, Seismic tomography and the assessment of uncertainty: Advances in Geophysics, **55**, 1–76.

Rawlinson, N., and Spakman, W., 2016, On the use of sensitivity tests in seismic tomography: Geophysical Journal International, **205**, 1221–1243.

Ren, H., Wu, R., and Wang, H., 2011, Wave equation least square imaging using the local angular hessian for amplitude correction: Geophysical Prospecting, **59**, 651–661.

Rickers, F., Fichtner, A., and Trampert, J., 2013, The Iceland-Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion: Earth and Planetary Science Letters, **367**, 39–51.

Sacchi, M. D., Wang, J., and Kuehl, H., 2007, Estimation of the diagonal of the migration blurring kernel through a stochastic approximation: SEG Expanded Abstracts, 2437–2441.

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

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

Spakman, W., 1991, Delay-time tomography of the upper mantle below Europe, the Mediterranean and Asia minor: Geophysical Journal International, **107**, 309–332.

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

Tang, Y., and Lee, S., 2015, Multi-parameter full wavefield inversion using non-stationary point-spread functions: SEG Technical Program Expanded Abstracts, 1111–1115.

Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, 1893–1903.

Tarantola, A., 2005, Inverse Problem Theory and Methods for Model Parameter Estimation: Society of Industrial and Appplied Mathematics.

Trampert, J., Fichtner, A., and Ritsema, J., 2013, Resolution tests revisited: the power of random numbers: Geophysical Journal International, **192**, 676–680.

Tromp, J., Tape, C., and Liu, Q., 2005, Seismic tomography, adjoint methods, time reversal, and banana-doughnut kernels: Geophysical Journal International, **160**, 195–216.

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

Valenciano, A. A., Biondi, B., and Guitton, A., 2006, Target-oriented wave-equation inversion. geophysics: Geophysics, **71**, A35–A38.

Wang, T., and Cheng, J. B., 2017, Elastic full-waveform inversion based on mode decomposition: the approach and mechanism: Geophysical Journal International, **209**, 606–622.

Wang, Y., Dong, L., Liu, Y., and Yang, J., 2016, 2D frequency-domain elastic full-waveform inversion using the block-diagonal pseudo-Hessian approximation: Geophysics, **81**, R247–R259.

Yang, J., Liu, Y., and Dong, L., 2016, Simultaneous estimation of velocity and density in acoustic multiparameter full-waveform inversion using an improved scattering-integral approach: Geophysics, **81**, R399–R415.

Yuan, Y. O., and Simons, F. J., 2014, Multiscale adjoint waveform-difference tomography using wavelets: Geophysics, **79**, WA79–WA95.

Yuan, Y. O., Simons, F. J., and Bozdağ, E., 2015, Multiscale adjoint waveform tomography for surface and body waves: Geophysics, **80**, R281–R302.

Zhu, H., Bozdag, E., and Tromp, J., 2015, Seismic structure of the european upper mantle based on adjoint tomography: Geophysical Journal Interational, **201**, 18–52.

Zhu, H., and Fomel, S., 2016, Building good starting models for full-waveform inversion using adaptive matching filtering misfit: Geophysics, **81**, U61–U72.

Zhu, H., Li, S., Fomel, S., Stadler, G., and Ghattas, O., 2016, A Bayesian approach to estimate uncertainty for full-waveform inversion using a prior information from depth migration: Geophysics, **81**, R307–R323.

Zhu, H., Luo, Y., Nissen-Meyer, T., Morency, C., and Tromp, J., 2009, Elastic imaging and time-lapse migration based on adjoint methods: Geophysics, **74**, WCA167–WCA177.