

Subspace Methods in Multi-Parameter Seismic Full Waveform Inversion

Yu Geng^{1,*}, Kristopher A. Innanen¹ and Wenyong Pan^{2,3}

¹ Department of Geosciences, CREWES Project, University of Calgary, Alberta T2N 1N4, Canada.

² Department of Geoscience, CREWES Project, University of Calgary, AB T2N 1N4, Canada.

³ Los Alamos National Laboratory, Geophysics Group, MS D452, Los Alamos, NM 87545, USA.

Received 2 April 2018; Accepted 6 March 2019

Abstract. In full waveform inversion (FWI) high-resolution subsurface model parameters are sought. FWI is normally treated as a nonlinear least-squares inverse problem, in which the minimum of the corresponding misfit function is found by updating the model parameters. When multiple elastic or acoustic properties are solved for, simple gradient methods tend to confuse parameter classes. This is referred to as parameter cross-talk; it leads to incorrect model solutions, poor convergence and strong dependence on the scaling of the different parameter types. Determining step lengths in a subspace domain, rather than directly in terms of gradients of different parameters, is a potentially valuable approach to address this problem. The particular subspace used can be defined over a span of different sets of data or different parameter classes, provided it involves a small number of vectors compared to those contained in the whole model space. In a subspace method, the basis vectors are defined first, and a local minimum is found in the space spanned by these. We examine the application of the subspace method within acoustic FWI in determining simultaneously updates for velocity and density. We first discuss the choice of basis vectors to construct the spanned space, from linear updates by distinguishing only the contributions of different parameter classes towards nonlinear updates by adding the contributions of higher-order perturbations of each parameter class. The numerical character of FWI solutions generated via subspace methods involving different basis vectors is then analyzed and compared with traditional FWI methods. The subspace methods can provide better reconstructions of the model, especially for the velocity, as well as improved convergence rates, while the computational costs are still comparable with the traditional FWI methods.

AMS subject classifications: 35L05, 35R30, 86A22

Key words: Waveform inversion, inverse problem, subspace method.

*Corresponding author. *Email addresses:* gengfish@gmail.com (Y. Geng), k.innanen@ucalgary.ca (K. A. Innanen), wenyongp@lanl.gov (W. Pan)

1 Introduction

In full waveform inversion (FWI) [11, 25, 27], subsurface model parameters are determined by minimizing an objective function measuring the difference between predicted data and recorded data, related through forward modelling. The forward modelling involves wave propagation physics which can range from scalar acoustic, to acoustic, up to viscoelastic anisotropic approximations, and beyond. Simultaneous inversion for different parameter classes (see, e.g., [3, 6, 16, 17, 19–21]), including for instance P-wave and S-wave velocities, density, as well as the various attenuation and anisotropic parameters, etc., are critical for a wide application of FWI in reservoir characterization. Similar to the mono-parameter inversion under the scalar acoustic approximation, in which only P-wave velocity is considered, in multi-parameter FWI a misfit function is set up to describe the distance between the recorded data and the predicted data, and FWI is treated as a nonlinear least squares problem, which can be solved by gradient-based methods or Newton-type methods. Multi-parameter inversion is more complicated than mono-parameter inversion, because the additional parameter classes increase the ill-posedness and the nonlinearity of the inverse problem. Different parameter classes can be more or less coupled, and it may be difficult to distinguish the contribution of each parameter class to changes in the data. Mitigating cross-talk between different parameter classes is a key issue. Studies have shown that the Hessian operator contains some information concerning the coupling between different parameter classes. Different ways of incorporating the inverse of the Hessian operator, especially in multi-parameter inversion, have been proposed to better decouple different parameter classes in the inversion. These include preconditioning the gradient using the pseudo Hessian matrix [24], quasi-Newton method, truncated Newton method [12–15, 18] and so on. Hierarchical strategies can be applied to successively invert different parameter classes to mitigate the ill-posedness of FWI [1, 2, 9]. In most cases involving incorporation of Hessian information, significant increases in computational cost ensue.

In both gradient-based methods and Newton-type methods (see, e.g., [7, 12, 13, 18, 22, 27, 31]), a line search scaling the descent direction tends to be necessary for convergence. One scalar is found for all parameter classes regardless of their contributions to the data. To combat cross-talk, distinguishing between the contributions of each parameter class during updating could be helpful in multi-parameter inversion.

Application of subspace methods in large-scale inverse problems was first discussed by [10, 23] as an approach to adjusting the update descent directions according to different parameter classes' contributions. Baumstein [5] showed that using an extended subspace method in multi-parameter inversion can also help to mitigate cross-talk. In subspace FWI, basis vectors are determined first, and the optimization problem is then solved in this spanned space to minimize the quadratic approximation of the misfit function, with only a few coefficients to be determined as compared to the traditional gradient-based or Newton-type methods. Although projection of the full Hessian or Gauss-Newton Hessian onto the subspace is needed for each iteration, the calculation is much cheaper com-

pared to Newton-type methods.

In this study, we evaluate different basis vectors, constructed from the gradients of two different acoustic parameter classes, and related Hessian-vector products (see, e.g., [12, 18]), for their ability to construct better multi-parameter inversions. Our results are restricted to the acoustic approximation (with varying density) both for forward modelling and inversion; however we expect our results to be by-and-large true for other parameterizations and elastic extensions.

2 Full waveform inversion and subspace methods: review

We use the frequency-space domain acoustic wave equation to describe wave motion:

$$\frac{\omega^2}{\rho(\mathbf{x})v^2(\mathbf{x})}u(\mathbf{x},\mathbf{x}_s,\omega) + \nabla \cdot \left(\frac{1}{\rho(\mathbf{x})} \nabla u(\mathbf{x},\mathbf{x}_s,\omega) \right) = f_s(\omega)\delta(\mathbf{x} - \mathbf{x}_s), \quad (2.1)$$

where ρ is the density and v is the velocity. Radiation patterns, or scattering patterns (see, e.g., [3, 19, 26, 28]) of v and ρ are plotted in Fig. 1. Compared to the radiation pattern of v , which is isotropic, the radiation pattern of ρ decreases progressively from small scattering angles, which are related to backward scattering, to large scattering angles, which are related to forward scattering (see, e.g., [16, 27]). This is consistent with the fact that a change of velocity affects the waves at all scattering angles, while a change in density only has significant influence on the amplitudes of the precritical reflections at small scattering angles. This is indicative of the possible difficulties of the reconstruction of these two parameters, especially using only a small range of near-offset traces.

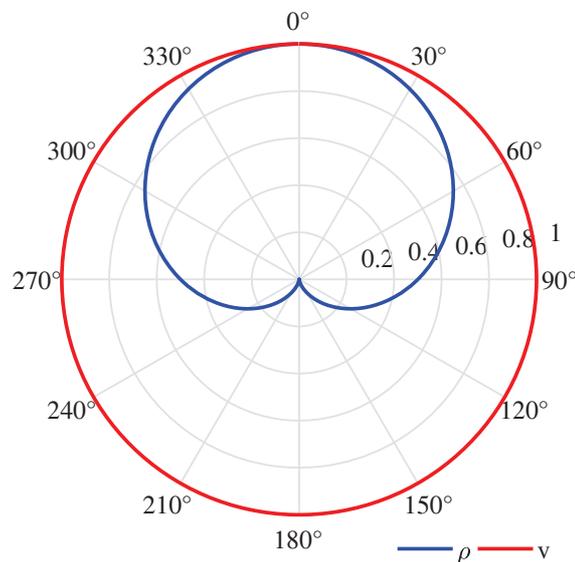


Figure 1: Radiation patterns of velocity and density in acoustic FWI.

Expressing model parameters of all types in one vector \mathbf{m} , the source term as $\mathbf{f}(\mathbf{x}_s, \omega)$ and the forward modelling operator, which is the discretized impedance matrix, as $\mathbf{F}(\mathbf{m}, \omega)$, a discretized form of the wave equation can be written in matrix form as

$$\mathbf{F}(\mathbf{m}, \omega) \mathbf{u}(\mathbf{m}, \mathbf{x}_s, \omega) = \mathbf{f}(\mathbf{x}_s, \omega), \quad (2.2)$$

The forward modeling (2.2) describes a nonlinear relationship between the wavefield $\mathbf{u}(\mathbf{m}, \mathbf{x}_s, \omega)$ and the model parameters \mathbf{m} .

In FWI inversion is formulated as an optimization problem, in which a model \mathbf{m} is sought which minimizes the misfit functional $\phi(\mathbf{m})$

$$\phi(\mathbf{m}) = \frac{1}{2} \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_g} \sum_{\omega} \|\mathbf{d}_{obs}(\mathbf{x}_s, \mathbf{x}_g, \omega) - \mathbf{d}_{syn}(\mathbf{m}, \mathbf{x}_s, \mathbf{x}_g, \omega)\|^2 = \frac{1}{2} \sum_{\mathbf{x}_s} \sum_{\mathbf{x}_g} \sum_{\omega} \|\Delta \mathbf{d}(\mathbf{m}, \mathbf{x}_s, \mathbf{x}_g, \omega)\|^2, \quad (2.3)$$

where $\mathbf{d}_{obs}(\mathbf{x}_s, \mathbf{x}_g, \omega)$ are the observed data at receiver location \mathbf{x}_g for each source location \mathbf{x}_s for one frequency ω , and $\mathbf{d}_{syn}(\mathbf{m}, \mathbf{x}_s, \mathbf{x}_g, \omega) = \mathbf{R}\mathbf{u}(\mathbf{m}, \mathbf{x}_s, \omega)$ are the synthetic data generated by simulating wave propagation in the current model iterate \mathbf{m} via (2.2), and sampling the wavefield with an operator \mathbf{R} to the receiver locations. $\Delta \mathbf{d}(\mathbf{m}, \mathbf{x}_s, \mathbf{x}_g, \omega)$ are the data residuals, which are defined as the difference between the observed and synthetic data.

Expanding the misfit functional (2.3) up to second order around the vicinity of the model \mathbf{m} ,

$$\phi(\mathbf{m} + \delta \mathbf{m}) = \phi(\mathbf{m}) + \langle \mathbf{g}, \delta \mathbf{m} \rangle + \frac{1}{2} \langle \mathbf{H} \delta \mathbf{m}, \delta \mathbf{m} \rangle + \mathcal{O}(\|\delta \mathbf{m}\|^3), \quad (2.4)$$

where \mathbf{g} and \mathbf{H} are the gradient and the Hessian operator of the misfit function, respectively. Since the misfit functional is generally non-quadratic, Eq. (2.4) represents a local quadratic approximation, in which the model can be iteratively updated:

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \alpha_n \delta \mathbf{m}. \quad (2.5)$$

The perturbation can be determined by a descent direction, which is the negative of the gradient in gradient-based methods, or it can be the solution of a Newton system, respectively

$$\begin{aligned} \delta \mathbf{m} &= -\mathbf{g}, \quad \text{or} \\ \delta \mathbf{m} &= -\mathbf{H}^{-1} \mathbf{g}. \end{aligned} \quad (2.6)$$

In either case, all parameters are updated in their specific descent directions scaled by a step-length α as in (2.5). This can be understood as a 1D subspace scheme, in which the optimization of the misfit functional in the complete model space is replaced by a 1D optimization of ϕ in the descent direction. The step-length is a constant for each parameter type; the update in each parameter is thus governed by the properties of the overall descent direction at each iteration, rather than the specific features of its own direction.

Subspace methods can be used for single parameter FWI problems. For instance, when the gradient associated with the current iteration and previous step information are combined, as in conjugate gradient methods and limited memory quasi-Newton method, a subspace of some lower dimension M is constructed within which the update vector is found. However, this induces a trade-off between an increasing computational cost per iteration and the possible decrease in number of iterations. In these methods, search directions are defined within a lower dimensional space (at least 2), as compared to the steepest descent method, which searches a 1D step-length in the full space (see, e.g., [32, 33]).

In the case of multi-parameter FWI, different datasets can be used to construct a subspace, and model space can also be divided into subspace. In this study, we consider the partitioning of the gradient into contributions from each parameter. We analyze frequency domain FWI in the acoustic approximation for both velocity and density, examining different subspaces.

Suppose that a model perturbation can be expressed in a space spanned by n basis vectors $\{\mathbf{a}_j\}$ as

$$\delta \mathbf{m} = \sum_{j=1}^n \mu_j \mathbf{a}_j. \quad (2.7)$$

Substituting Eq. (2.7) into the expansion of the misfit functional in (2.4) we obtain

$$\phi(\mathbf{m} + \delta \mathbf{m}) = \phi(\mathbf{m}) + \sum_{j=1}^n \mu_j \langle \mathbf{g}, \mathbf{a}_j \rangle + \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \mu_j \mu_i \langle \mathbf{H} \mathbf{a}_j, \mathbf{a}_i \rangle + \dots \quad (2.8)$$

The coefficient μ_j for each basis vector \mathbf{a}_j can be found using

$$\mu_j \langle \mathbf{g}, \mathbf{a}_j \rangle + \sum_{i=1}^n \mu_j \langle \mathbf{H} \mathbf{a}_j, \mathbf{a}_i \rangle \approx 0. \quad (2.9)$$

Written in matrix form

$$\mathbf{A}^T \mathbf{g} + \mathbf{A}^T \mathbf{H} \mathbf{A} \boldsymbol{\mu} \approx 0. \quad (2.10)$$

The coefficients $\boldsymbol{\mu}$ can then be determined from the projection of the gradient and the Hessian onto the subspace in the form

$$\boldsymbol{\mu} = - \left(\mathbf{A}^T \mathbf{H} \mathbf{A} \right)^{-1} \mathbf{A}^T \mathbf{g}. \quad (2.11)$$

Since the subspace is only n dimensions, $\mathbf{A}^T \mathbf{H} \mathbf{A}$ is an $n \times n$ matrix and is generally simple to invert. When the term related to the second-order partial derivative of the data respect to the model parameters in the Hessian can be neglected, the Gauss-Newton Hessian can be used in the equation to evaluate the coefficient.

3 Choice of subspace basis vectors

3.1 Subspace basis vectors for linear updates

3.1.1 2D choice

We could straightforwardly choose the descent direction (e.g., the negative of the gradient) of each parameter as our subspace basis vectors. In this case, the descent direction vector for each parameter is simply extended into the whole model space. For instance, in the case of two parameters v and ρ , the basis vectors \mathbf{a}_i are the extension the gradients of each parameter \mathbf{g}_v and \mathbf{g}_ρ in the whole model space:

$$\mathbf{a}_1 = \begin{bmatrix} -\mathbf{g}_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ -\mathbf{g}_\rho \end{bmatrix}. \quad (3.1)$$

In multiparameter FWI, when the position of model parameters are used before the parameter class to sort the elements in the Hessian matrix, the Hessian matrix has a block structure with diagonal blocks \mathbf{H}_{vv} and $\mathbf{H}_{\rho\rho}$, and off-diagonal blocks $\mathbf{H}_{v\rho}$ and $\mathbf{H}_{\rho v}$. The Hessian matrix can be written as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{vv} & \mathbf{H}_{v\rho} \\ \mathbf{H}_{\rho v} & \mathbf{H}_{\rho\rho} \end{bmatrix} \quad (3.2)$$

with coefficients μ given by

$$\mu = \begin{bmatrix} \mathbf{g}_v^T \mathbf{H}_{vv} \mathbf{g}_v & \mathbf{g}_v^T \mathbf{H}_{v\rho} \mathbf{g}_\rho \\ \mathbf{g}_\rho^T \mathbf{H}_{\rho v} \mathbf{g}_v & \mathbf{g}_\rho^T \mathbf{H}_{\rho\rho} \mathbf{g}_\rho \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}_v^T \mathbf{g}_v \\ \mathbf{g}_\rho^T \mathbf{g}_\rho \end{bmatrix} \quad (3.3)$$

which, because it involves only Hessian-vector products (as in Hessian-free Newton methods (see, e.g., [12, 18])), is easily calculated. In this case, two Hessian-vector products are needed for each calculation of the coefficient, and a 2×2 matrix is inverted. Compared to steepest descent methods, which involves a line search, and conjugate gradient methods for linear problems, which within the inner loop of the truncated Gauss-Newton method involves the calculation

$$\mu = \frac{\langle \mathbf{g}, \mathbf{g} \rangle}{\langle \mathbf{H} \mathbf{g}, \mathbf{g} \rangle}, \quad (3.4)$$

we observe several differences. The step lengths for the updates in each parameter type are different, and the model is updated by the perturbation for each parameter class as

$$\delta v = -\mu_v \mathbf{g}_v, \quad \delta \rho = -\mu_\rho \mathbf{g}_\rho. \quad (3.5)$$

3.1.2 4D choice

These basis vectors can be extended in ways which can ultimately be designed to prevent leakage between gradient components. For instance, we can define

$$\mathbf{a}_1 = \begin{bmatrix} -\mathbf{g}_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} -\mathbf{g}_\rho \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 \\ -\mathbf{g}_v \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 \\ -\mathbf{g}_\rho \end{bmatrix}. \quad (3.6)$$

In this case,

$$\boldsymbol{\mu} = \begin{bmatrix} \mathbf{g}_v^T \mathbf{H}_{vv} \mathbf{g}_v & \mathbf{g}_v^T \mathbf{H}_{v\rho} \mathbf{g}_\rho & \mathbf{g}_v^T \mathbf{H}_{v\rho} \mathbf{g}_v & \mathbf{g}_v^T \mathbf{H}_{v\rho} \mathbf{g}_\rho \\ \mathbf{g}_\rho^T \mathbf{H}_{vv} \mathbf{g}_v & \mathbf{g}_\rho^T \mathbf{H}_{v\rho} \mathbf{g}_\rho & \mathbf{g}_\rho^T \mathbf{H}_{v\rho} \mathbf{g}_v & \mathbf{g}_\rho^T \mathbf{H}_{v\rho} \mathbf{g}_\rho \\ \mathbf{g}_v^T \mathbf{H}_{\rho v} \mathbf{g}_v & \mathbf{g}_v^T \mathbf{H}_{\rho v} \mathbf{g}_\rho & \mathbf{g}_v^T \mathbf{H}_{\rho\rho} \mathbf{g}_v & \mathbf{g}_v^T \mathbf{H}_{\rho\rho} \mathbf{g}_\rho \\ \mathbf{g}_\rho^T \mathbf{H}_{\rho v} \mathbf{g}_v & \mathbf{g}_\rho^T \mathbf{H}_{\rho v} \mathbf{g}_\rho & \mathbf{g}_\rho^T \mathbf{H}_{\rho\rho} \mathbf{g}_v & \mathbf{g}_\rho^T \mathbf{H}_{\rho\rho} \mathbf{g}_\rho \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}_v^T \mathbf{g}_v \\ \mathbf{g}_\rho^T \mathbf{g}_v \\ \mathbf{g}_v^T \mathbf{g}_\rho \\ \mathbf{g}_\rho^T \mathbf{g}_\rho \end{bmatrix}. \quad (3.7)$$

Here four Hessian-vector products are required and a 4×4 matrix is inverted to calculate the coefficients for four basis vectors. We note that this represents an increase in computational cost, and so any increase in cross-talk suppression so derived is not “free”, though it may be quite inexpensive.

3.1.3 6D choice

The rate of change of the descent vectors can also be used to construct basis vectors, as discussed by [10], e.g.,

$$\begin{aligned} \mathbf{a}_1 &= \begin{bmatrix} -\mathbf{g}_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ -\mathbf{g}_\rho \end{bmatrix}, \\ \mathbf{a}_3 &= \begin{bmatrix} \mathbf{H}_{vv} & \mathbf{H}_{v\rho} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{g}_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 & 0 \\ \mathbf{H}_{\rho v} & \mathbf{H}_{\rho\rho} \end{bmatrix} \begin{bmatrix} \mathbf{g}_v \\ 0 \end{bmatrix}, \\ \mathbf{a}_5 &= \begin{bmatrix} \mathbf{H}_{vv} & \mathbf{H}_{v\rho} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{g}_\rho \end{bmatrix}, \quad \mathbf{a}_6 = \begin{bmatrix} 0 & 0 \\ \mathbf{H}_{\rho v} & \mathbf{H}_{\rho\rho} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{g}_\rho \end{bmatrix}. \end{aligned} \quad (3.8)$$

However, as pointed out in Baumstein’s study [5], when the misfit function is not locally quadratic, curvature information obtained from the Hessian may be far from that of the global minimum, and the coefficients so obtained may not lead to an improved search direction. Also, when using the Hessian in the construction of the basis vectors, the computational cost will tend upward, possibly dramatically.

3.2 Subspace basis vectors leading to nonlinear updates

When considering the second-order scattering in the Hessian operator as corrections to the gradient (see, e.g., [4,7,22,30]), higher-order perturbations (see, e.g., [29]) can be used

as the basis vectors in addition to the gradients for each parameter type. In which case, not only the contributions of different parameter classes, but also the contributions of higher-order perturbations from each parameter class, can be distinguished, with additional nonlinearity being added to each parameter class during each update. The full Hessian can be decomposed as

$$\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2, \quad (3.9)$$

where \mathbf{H}_1 contains the second-order partial derivative of the data with respect to the model parameters, and \mathbf{H}_2 is the Gauss-Newton Hessian operator. Under this assumption, the model perturbation can be modified from (2.5) to

$$\delta \mathbf{m} = -\mathbf{H}_2^{-1} \left(\mathbf{g} - \mathbf{H}_1 \mathbf{H}_2^{-1} \mathbf{g} \right). \quad (3.10)$$

To calculate this model perturbation, the inverse of the Gauss-Newton Hessian \mathbf{H}_2^{-1} is needed. To avoid the high cost of direct calculation of this term, a quasi-Newton method or truncated Newton method can be used. Suppose that the $\delta \mathbf{m}_1 = -\mathbf{H}_2^{-1} \mathbf{g}$ is the perturbation obtained using a linearized inversion, e.g., a truncated Gauss-Newton method, and also ignore the affect of the inverse of the Gauss-Newton Hessian \mathbf{H}_2^{-1} outside the bracket. The basis vectors can be a combination between the gradient vectors for each parameter class and also its related higher-order perturbations:

$$\mathbf{a}_1 = \begin{bmatrix} -\mathbf{g}_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} (\mathbf{H}_1 \delta \mathbf{m}_1)_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 \\ -\mathbf{g}_\rho \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 \\ (\mathbf{H}_1 \delta \mathbf{m}_1)_\rho \end{bmatrix}. \quad (3.11)$$

Compared to the 4D subspace constructed from the gradient vectors (3.6), one more Hessian-vector $\mathbf{H}_1 \delta \mathbf{m}_1$ is needed during the calculation of the coefficients.

To find the update using the subspace basis in Eq. (3.11), a perturbation model $\delta \mathbf{m}_1$ is needed to form the basis vectors \mathbf{a}_2 and \mathbf{a}_4 . The details of the calculation of this perturbation model is critical to the whole nonlinear update, since cross-talk artifacts should be removed by the inverse of Gauss-Newton Hessian operator when calculating this term. However, the effect of the inverse of the Gauss-Newton Hessian operator outside the bracket in (3.10) may not be ignored since cross-talk artefacts may still exist in the model updates founded within subspace basis vectors (3.11). Therefore, directly using perturbation model $\delta \mathbf{m}_1$ to construct the basis vectors could be a better choice to involve non-linearity of each parameter class into each update

$$\begin{aligned} \mathbf{a}_1 &= \begin{bmatrix} (\delta \mathbf{m}_1)_v \\ 0 \end{bmatrix}, & \mathbf{a}_2 &= \begin{bmatrix} -\mathbf{H}_2^{-1} (\mathbf{H}_1 \delta \mathbf{m}_1)_v \\ 0 \end{bmatrix}, \\ \mathbf{a}_3 &= \begin{bmatrix} 0 \\ (\delta \mathbf{m}_1)_\rho \end{bmatrix}, & \mathbf{a}_4 &= \begin{bmatrix} 0 \\ -\mathbf{H}_2^{-1} (\mathbf{H}_1 \delta \mathbf{m}_1)_\rho \end{bmatrix}. \end{aligned} \quad (3.12)$$

In this study, we use an l-BFGS approach to approximate the inverse of the Hessian \mathbf{H}_2^{-1} as a preconditioner for the descent direction to calculate the perturbation model

$\delta\mathbf{m}_1$ in (3.11); a pseudo Hessian can also be used with the assumption that in the multi-parameter case, each block in the Hessian for different parameter classes is roughly diagonal [8, 14, 24]. We use truncated Gauss-Newton method to calculate the perturbation model $\delta\mathbf{m}_1$ in (3.12), while l-BFGS approach is used to approximate the inverse of the Hessian \mathbf{H}_2^{-1} as a preconditioner when needed. In cases where the basis vectors may have cross-coupling with each other, when for instance we use gradients of different parameter classes (3.6), or when we use the Hessian to construct the basis vectors for each parameter class as in (3.8) and (3.11), orthogonalizing the basis vectors may be needed to avoid linear dependence:

$$\hat{\mathbf{a}}_2 = \mathbf{a}_2 - \frac{\mathbf{a}_1 \mathbf{a}_2}{\mathbf{a}_1 \mathbf{a}_1} \mathbf{a}_1. \quad (3.13)$$

4 Numerical examples

We test the application of the subspace method in both gradient-type and Newton-type simultaneous FWI updates of velocity and density, in both transmission-type and reflection-type scenarios. Steepest descent and truncated Newton methods are used, but the line search components are replaced with the subspace methods discussed in the previous section.

4.1 Subspace methods in multi-parameter transmission FWI

4.1.1 The subspace method as a substitute for the line search

Fig. 2 shows true and initial models for both velocity and density, and the model is 50×100 grid nodes in size, with grid interval of 10 m in each direction; the main structural components of the models are Gaussian balls. 49 sources are at the top of the model, with the first source located at $x_s = 20$ m, and 100 receivers are at the bottom of the model. 5 outer iterations are used for both the steepest descent and truncated Gauss-Newton methods, where 10 CG iterations are used to solve the linearized inverse problem inside the truncated Gauss-Newton methods. Three data frequencies, 8, 10 and 15 Hz, are used for the inversion and the Gauss-Newton Hessian is used. We test both 2D and 4D subspace methods. The inversion results are plotted in Fig. 3. In Fig. 4 horizontal profiles through the inverted velocity and density models at $z = 0.25$ km are plotted. To compare the convergence of the subspace FWI with the original FWI with line searching, we show both the misfit and model error versus iteration in Fig. 5.

From the inversion results, we observe that augmenting the steepest descent method with a subspace approach improves the update for the velocity with both 2D and 4D gradient based basis vectors. However, the updating of the density worsens due to density-velocity cross-talk. This is explainable through the scattering radiation patterns as shown in Fig. 1 generated by isotropic point inclusions: for the density the radiation pattern is strong at small scattering angles, which makes it difficult to update density

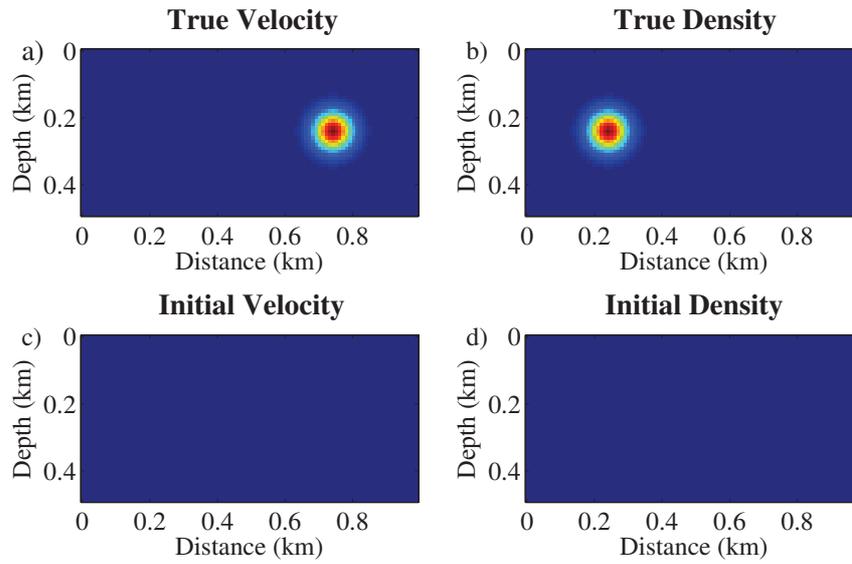
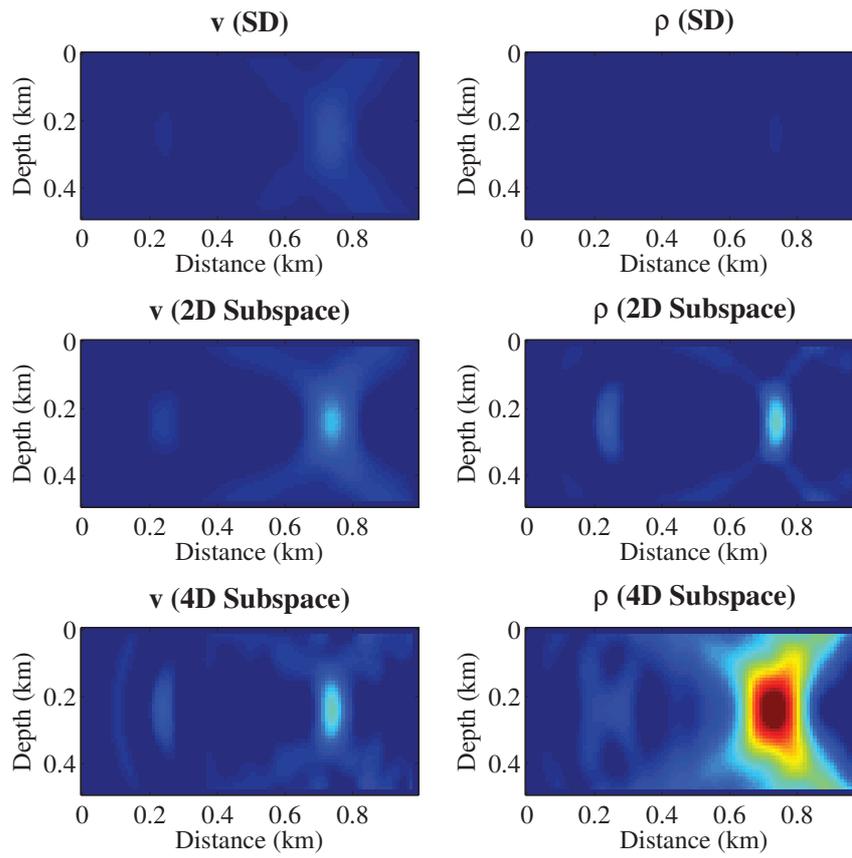


Figure 2: True velocity a) and density b). Initial velocity c) and density d).



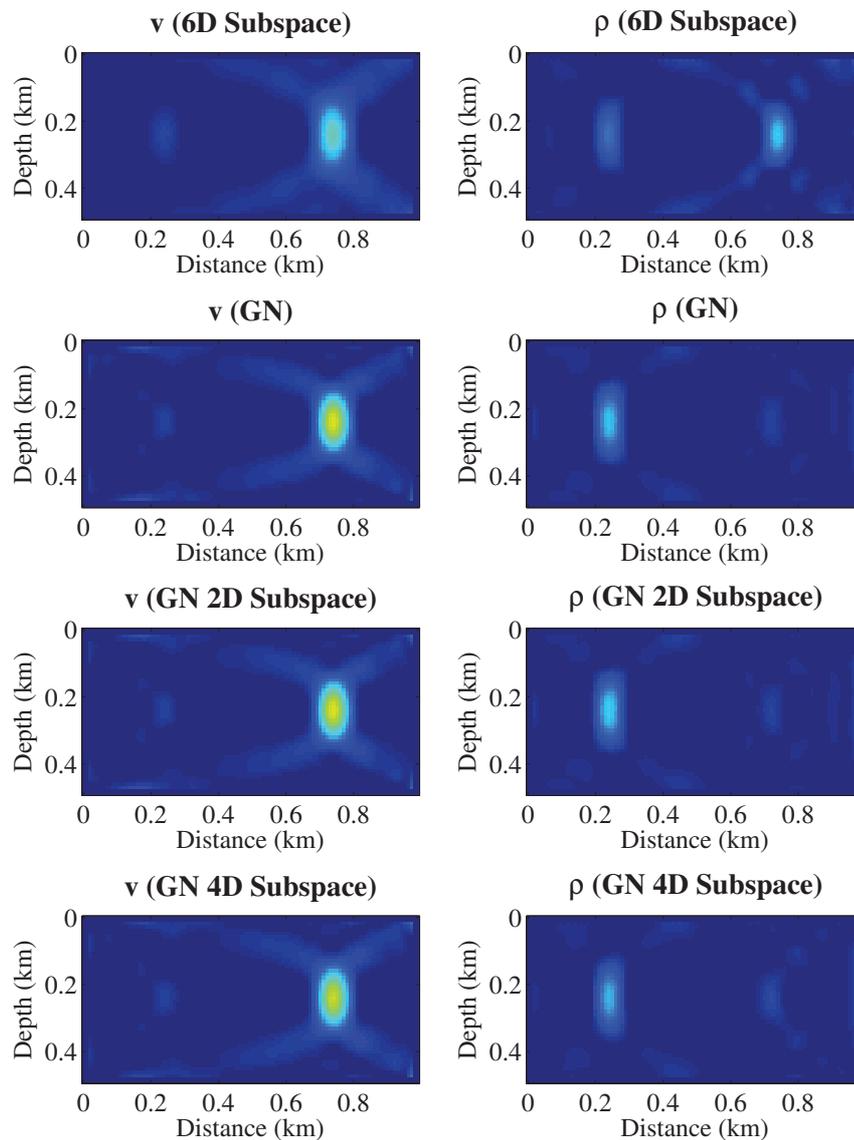


Figure 3: Inverted velocity and density using different methods. Here, SD stands for steepest descent method, GN stands for truncated Gauss-Newton method. Subspace method is used to replace line searching for each iteration.

in transmission experiments. This reflects the fact that recorded data are more sensitive to velocity variations, and barely respond to density variations. Cross talk between velocity and density is strong and positive-signed, which leads to the expectation that the 4D subspace method will not outperform the 2D method for any parameter, and in fact will tend to slow down the convergence. The steepest descent subspace method exhibits better convergence only for the first iteration as compared to the truncated Gauss-

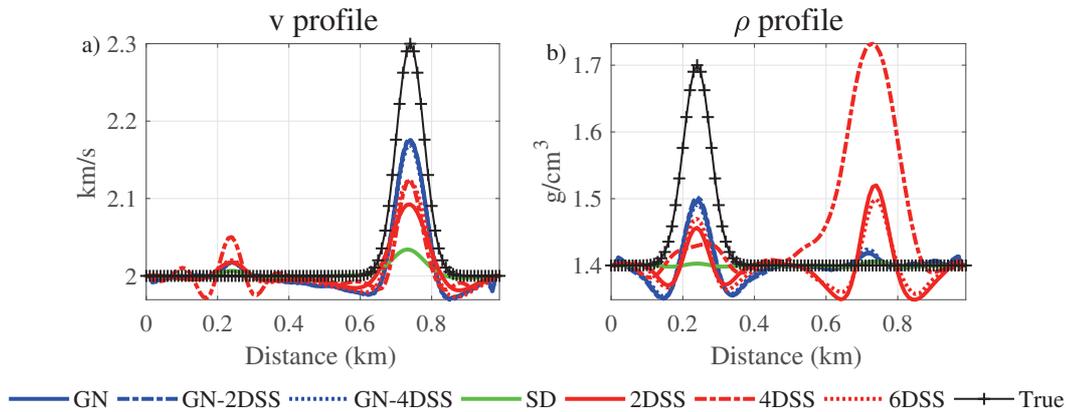


Figure 4: Profile for velocity and density at $z=0.25$ km. Here, SD stands for steepest descent method, GN stands for truncated Gauss-Newton method, and SS stands for subspace method.

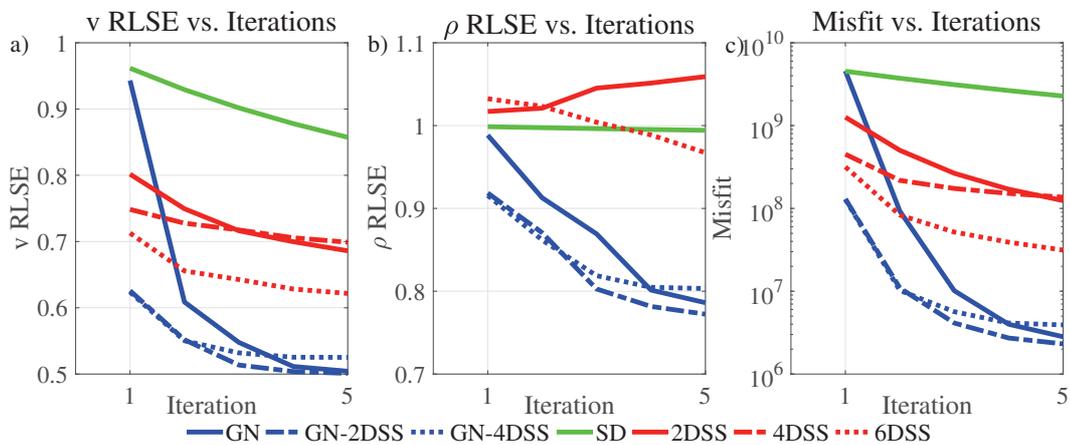


Figure 5: Comparison of the convergence of different methods with/without subspace method replacing line searching for each iteration. Here, SD stands for steepest descent method, GN stands for truncated Gauss-Newton method, and SS stands for subspace method.

Newton method. Rather, when the subspace method is combined with the truncated Gauss-Newton method, the cross-talk between velocity and density is reduced. Here both the 2D and 4D subspace methods provide improved inversion results, especially for density.

4.1.2 A subspace basis constructed from the nonlinear update

We next analyze the use of the higher-order perturbations as discussed in Section 3.2 to construct the basis vectors. A maximum of 10 inner iterations in truncated Newton and Gauss-Newton method and 20 outer iterations are used for the inversion. In Fig. 6 the results of different methods are compared. In Fig. 7 the corresponding model profiles for velocity and density are plotted at $z=0.25$ km. We observe:

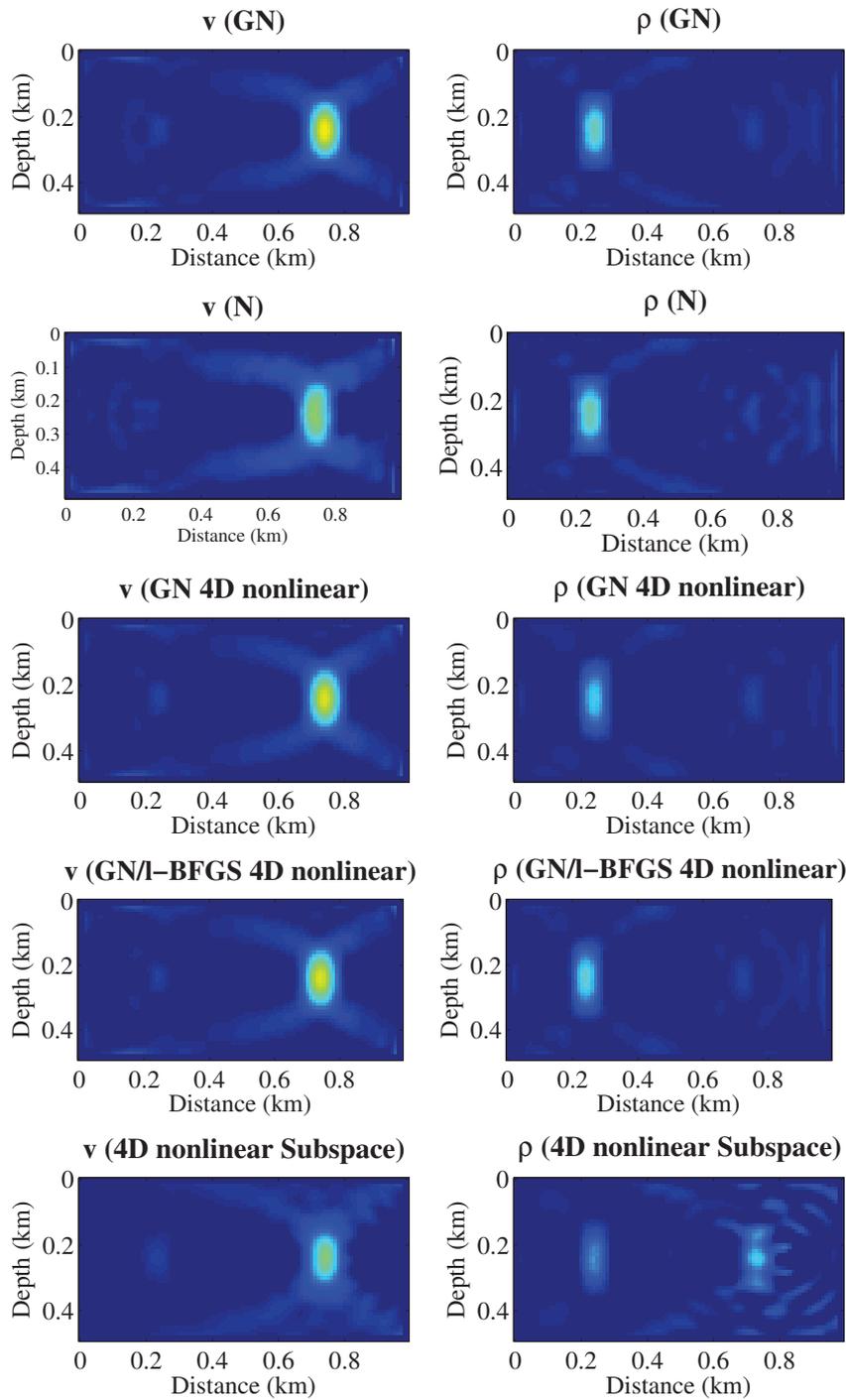


Figure 6: Inverted velocity and density using different methods. Here, GN stands for truncated Gauss-Newton method, N stands for truncated Newton method, nonlinear inversion using 4D subspace basis as in Eq. (3.12) with/without l-BFGS preconditioning are indicated by GN 4D nonlinear and GN/l-BFGS 4D nonlinear, and 4D nonlinear subspace stands for nonlinear inversion using 4D subspace basis as in Eq. (3.11).

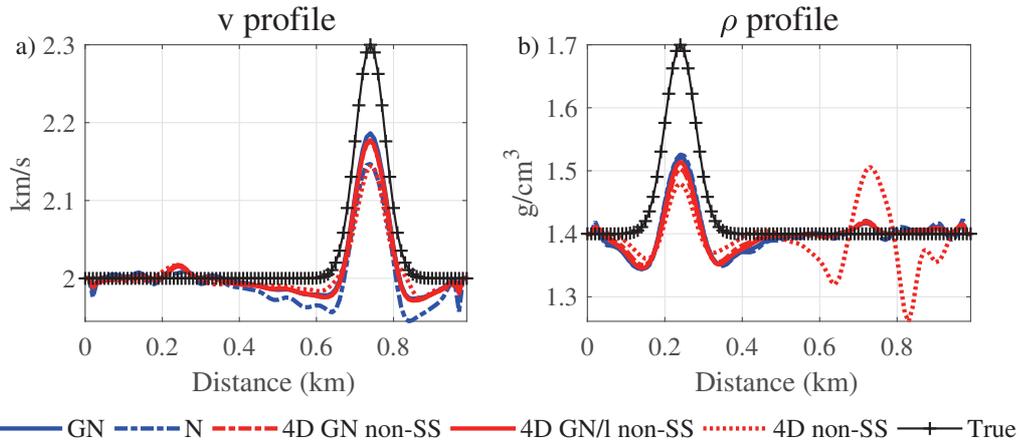


Figure 7: Profile for velocity and density at $z=0.25$ km. Here, GN stands for truncated Gauss-Newton method, N stands for truncated Newton method, nonlinear inversion using 4D subspace basis as in Eq. (3.12) with/without I-BFGS preconditioning are indicated by GN 4D nonlinear and GN/I-BFGS 4D nonlinear, and 4D nonlinear subspace stands for nonlinear inversion using 4D subspace basis as in Eq. (3.11).

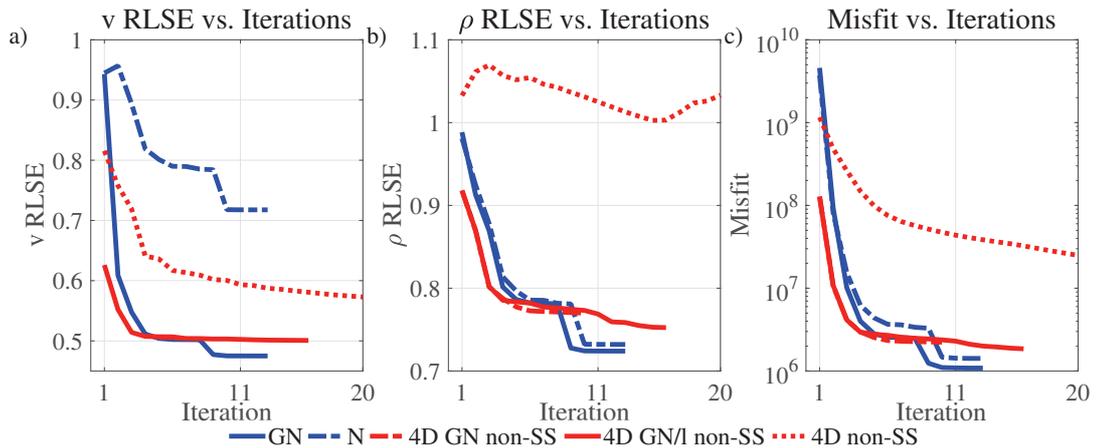


Figure 8: Comparison of the convergence of the truncated Newton method and nonlinear subspace method. Here, GN stands for truncated Gauss-Newton method, N stands for truncated Newton method, nonlinear inversion using 4D subspace basis as in Eq. (3.12) with/without I-BFGS preconditioning are indicated by GN 4D nonlinear and GN/I-BFGS 4D nonlinear, and 4D nonlinear subspace stands for nonlinear inversion using 4D subspace basis as in Eq. (3.11).

1. Using the full Hessian rather than the Gauss-Newton Hessian in the truncated Newton method does not appear to provide a better update, especially for velocity;
2. Using the model perturbation obtained from a Gauss-Newton update and its related higher-order perturbation term with or without I-BFGS pre-conditioning for the 4D Gauss-Newton nonlinear subspace updates, the inversion results converge faster in the early stage for both velocity and density; and

Table 1: Average computational costs of different FWI methods with/without subspace method for each iteration. Here, SD stands for steepest descent method, GN stands for truncated Gauss-Newton method, and SS stands for subspace method.

Methods	Maximum inner iteration	No. of Hessian-vector products for subspace	Average time (s)
SD	N/A	N/A	57.1
2D SS	N/A	2	41.6
4D SS	N/A	4	77.1
4D nonlinear SS	N/A	5	96.3
6D SS	N/A	6	112.9
GN	10	N/A	230.0
GN 2D SS	10	2	270.0
GN 4D SS	10	4	301.0
GN 4D nonlinear	10	5	336.2
GN/l-BFGS 4D nonlinear	10	5	338.1
truncated Newton	10	N/A	312.5

- Using the gradient and the higher-order perturbation term (i.e., the model perturbation approximated within the l-BFGS update) to construct the subspace basis vectors for the 4D nonlinear subspace updates, the velocity inversion is better than that obtained with the truncated Newton method, but (because the Gauss-Newton Hessian is replaced with the identity matrix), cross-talk artifacts strongly impact the density model.

4.1.3 Computational aspects

Computational costs of performing the different FWI methods with/without subspace method as used in Section 4.1.1 and Section 4.1.2 are averaged over number of iterations and listed in Table 1. Truncated Gauss-Newton/Newton methods have larger computational costs because of the computation of Hessian matrix in the inner iterations. With the computation of additional Hessian-vector products needed for the construction of subspace, increasing the number of subspace basis vectors also slightly increases the computational cost, however, it is still comparable with the original FWI methods.

4.2 Subspace methods in multi-parameter reflection FWI

The application of the subspace method to reflection-mode data is more complex. We take a small subregion of the original Marmousi-II model as our true model, and create an initial model by smoothing the true model with a Gaussian smoothing window under the water layer, as shown in Fig. 9. The new model is 81×161 grid nodes in size, with grid interval of 20 m in each direction. 4 frequencies (3, 5, 8, 12 Hz) are inverted simultaneously, and a maximum of 20 iterations are carried out in all the methods, where

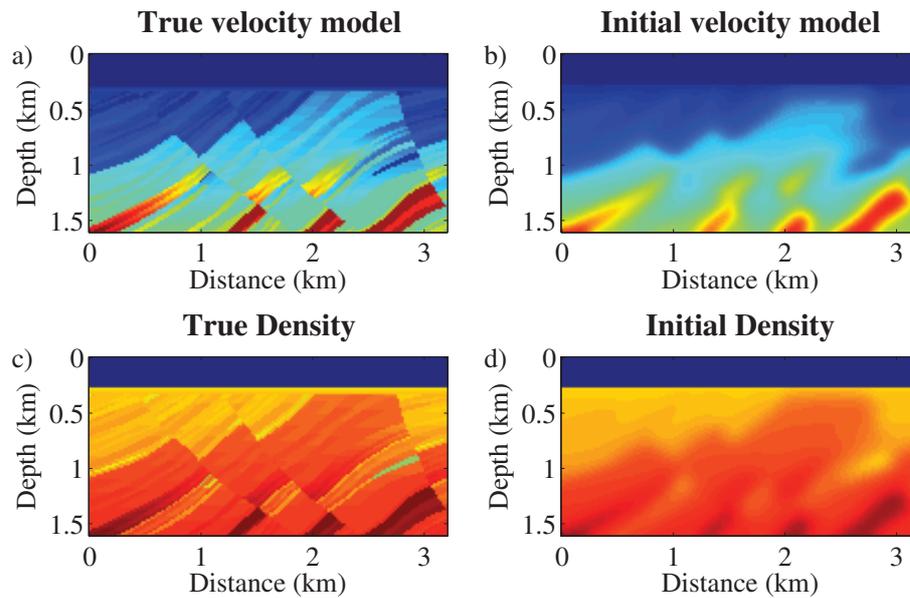


Figure 9: True velocity a) and density b). Initial velocity c) and density d).

10 inner iterations are used for the truncated Gauss-Newton type methods. Inversion results for each method under study (a 2D subspace method, a 6D subspace method, a truncated Gauss-Newton method, and a truncated Gauss-Newton method combined with a 2D subspace method) are plotted in Fig. 10. In Fig. 11 the convergence profiles of these four methods are plotted. We observe that using the subspace methods in place of line searches improves the velocity updates noticeably, but this can come at the cost of overestimation of the density.

5 Discussions

Using subspace methods in multi-parameter FWI in place of line search methods is observed to positively alter descent directions, by accounting for the varying contribution of each parameter class. Furthermore, higher-order perturbations can also be included to construct subspace basis vectors, so that the different contributions of higher-order perturbations for each parameter class can be included for each update and a nonlinear update for each parameter class can be constructed. The Hessian operator is involved in the calculation of subspace vector coefficients; compared to Newton-type methods, subspace methods appear to be able to better incorporate the contributions of the higher-order model perturbation in the descent direction for each parameter class. Therefore, subspace methods in multi-parameter FWI may be useful tools for mitigating cross-talk between parameter classes, it being capable of updating each parameter class with a tailored step length. Also, it can when desired include nonlinear/multiple-scattering re-

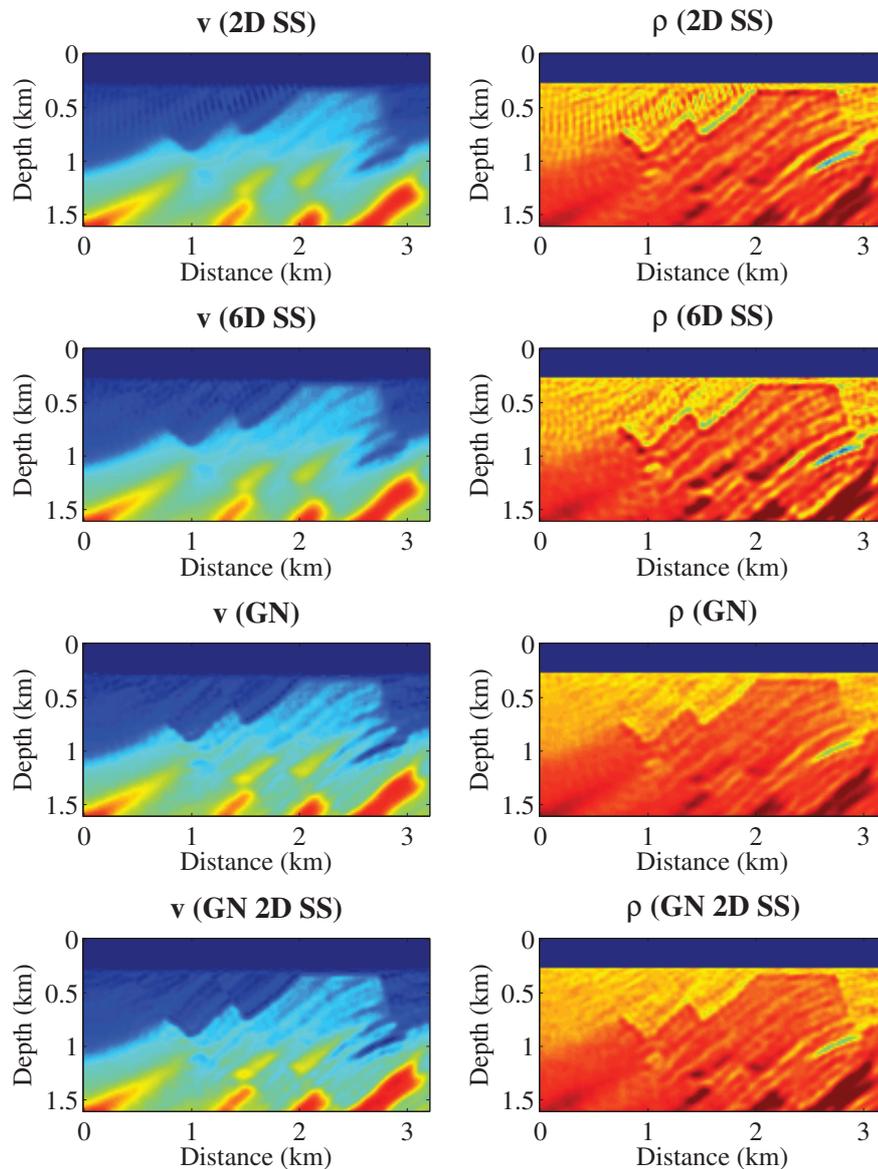


Figure 10: Inverted velocity and density using different methods. Results are plotted in the same scale of the true model. Here, GN stands for truncated Gauss-Newton method, and SS stands for subspace method.

lated perturbations as a correction to the single-scattering related perturbations for different parameter classes.

Advantages of the subspace method, then, are that

1. Gradients with different parameters are considered individually. A least-squares inversion is performed within the subspace, which is constructed including accounting for the individual influence of each parameter class;

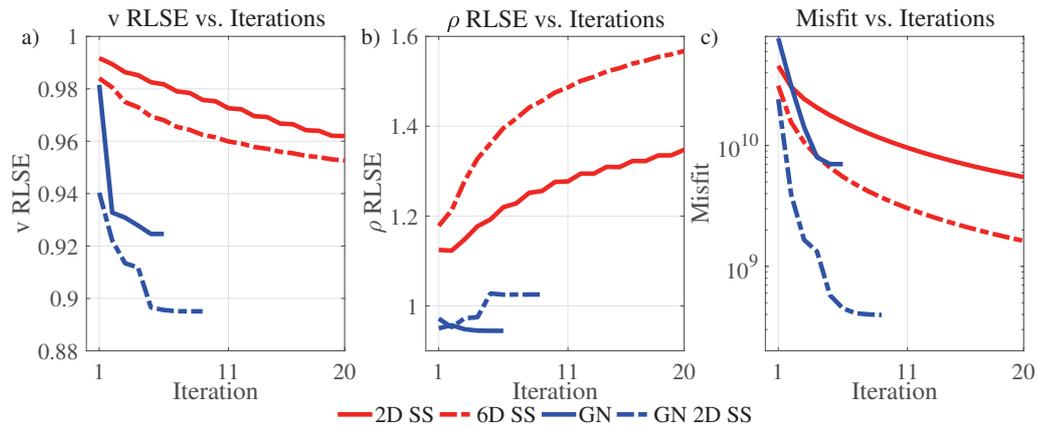


Figure 11: Comparison of the convergence of the truncated Newton method and nonlinear subspace method. Here, GN stands for truncated Gauss-Newton method, and SS stands for subspace method.

2. Coefficients are calculated using the Hessian-vector product, and is thus much cheaper than Newton-type methods (it is comparable to truncated Gauss-Newton/Newton methods);
3. The approach shows signs that they may be useful for attenuation of crosstalk in multi-parameter FWI; and
4. Nonlinear / multiple scattering related perturbations can be used to construct the basis vectors, so that the intrinsic nonlinearity of the FWI problem can be more directly handled.

However, disadvantages of the subspace method can also be significant, since

1. The theory relies on the assumption that the Hessian correctly captures the full behavior of the misfit function. When the misfit function is not locally quadratic, there is no guarantee that the global minimum can be found;
2. An approximate Hessian is usually used rather than the exact Hessian, which introduces difficult-to-quantify inaccuracies; and
3. It is possible that a model perturbation calculated from (2.11) will be too large for the misfit function to be considered approximately quadratic. In this case, another scaling factor may be needed to modify the coefficients so that updates converge.

6 Conclusions

We analyze the application of subspace methods to acoustic FWI and the simultaneous updating of velocity and density. Subspace methods can be used in place of the line

search in traditional implementations of FWI to obtain step lengths for different parameter classes. The Gauss-Newton Hessian in Hessian-vector products is involved to find the local minimum in the spanned space. We formulate different basis vectors, from gradient of each parameter class to higher-order perturbations related to each parameter class, to construct different spanned spaces to distinguish the contributions of different parameter classes for linear updates as well as the contributions of higher-order perturbations of each parameter class for nonlinear updates. The behavior of the subspace methods for both linear and nonlinear updates are compared with traditional FWI methods. The subspace methods exhibit improved convergence rates, as well as better reconstruction of the velocity model. The reconstruction of density model, however, remains affected by cross-talk artifacts, especially when the Hessian is not considered as in gradient-based methods.

Acknowledgments

We thank the associate editor Dr. Lianjie Huang, two anonymous reviewers for valuable suggestions and comments that ultimately helped to improve this paper. We thank the sponsors of CREWES for support. This work was funded by CREWES (Consortium for Research in Elastic Wave Exploration Seismology), and by NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13, and in part through the Canada First Research Excellence Fund.

References

- [1] T. Alkhalifah. Research note: Insights into the data dependency on anisotropy: an inversion prospective. *Geophysical Prospecting*, 64(2):505–513, 2016. 10.4208/cicp.OA-2018-008710.1111/1365-2478.12345.
URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/1365-2478.12345>.
- [2] T. Alkhalifah and M. Nabil. A new parameterization for waveform inversion in acoustic orthorhombic media. *Geophysics*, 81(4):R157–R171, 2016. ISSN 0016-8033. 10.4208/cicp.OA-2018-008710.1190/geo2015-0635.1.
- [3] T. Alkhalifah and R. E. Plessix. A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study. *Geophysics*, 79(3):R91–R101, 2014. ISSN 0016-8033. 10.4208/cicp.OA-2018-008710.1190/Geo2013-0366.1.
URL <Go to ISI>://WOS:000338322900029.
- [4] T. Alkhalifah and Z. D. Wu. Multiscattering inversion for low-model wavenumbers. *Geophysics*, 81(6):R417–R428, 2016. ISSN 0016-8033. 10.4208/cicp.OA-2018-008710.1190/Geo2015-0650.1. URL <Go to ISI>://WOS:000392926400022.
- [5] A. Baumstein. Extended subspace method for attenuation of crosstalk in multi-parameter full wavefield inversion. In *SEG Technical Program Expanded Abstracts*, pages 1121–1125, 2014.
- [6] A. Fichtner. *Full Seismic Waveform Modelling and Inversion*. Springer-Verlag Berlin Heidelberg, 2011. ISBN 978-3-642-15806-3.
- [7] Y. Geng, W. Pan, and K. A. Innanen. Frequency-domain full-waveform inversion with non-linear descent directions. *Geophysical Journal International*, 213(2):739–756, 2018.

- 10.4208/cicp.OA-2018-008710.1093/gji/ggy002.
URL <http://dx.doi.org/10.1093/gji/ggy002>.
- [8] K. A. Innanen. Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion. *Geophysical Journal International*, 199(2):717–734, 2014. ISSN 0956-540X. 10.4208/cicp.OA-2018-008710.1093/gji/ggu291.
URL <Go to ISI>://WOS:000343411700008.
- [9] W. Jeong, H. Y. Lee, and D. J. Min. Full waveform inversion strategy for density in the frequency domain. *Geophysical Journal International*, 188(3):1221–1242, 2012. ISSN 0956-540X. 10.4208/cicp.OA-2018-008710.1111/j.1365-246X.2011.05314.x.
URL <Go to ISI>://WOS:000300567400033.
- [10] B. L. N. Kennett, M. S. Sambridge, and P. R. Williamson. Subspace methods for large inverse problems with multiple parameter classes. *Geophysical Journal-Oxford*, 94(2):237–247, 1988. ISSN 0952-4592. 10.4208/cicp.OA-2018-0087DOI 10.1111/j.1365-246X.1988.tb05898.x.
URL <Go to ISI>://WOS:A1988P610200005.
- [11] P. Lailly. The seismic inverse problem as a sequence of before stack migrations. In *Conference on Inverse Scattering, Theory and Application, Society of Industrial and Applied Mathematics, Expanded Abstracts*, pages 206–220, 1983.
- [12] L. Métivier, R. Brossier, J. Virieux, and S. Operto. Full waveform inversion and the truncated Newton method. *Siam Journal on Scientific Computing*, 35(2):B401–B437, 2013. ISSN 1064-8275. 10.4208/cicp.OA-2018-008710.1137/120877854.
URL <Go to ISI>://WOS:000318404100035.
- [13] L. Métivier, F. Bretaudeau, R. Brossier, S. Operto, and J. Virieux. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures. *Geophysical Prospecting*, 62(6):1353–1375, 2014. ISSN 0016-8025. 10.4208/cicp.OA-2018-008710.1111/1365-2478.12136. URL <Go to ISI>://WOS:000344222300013.
- [14] L. Métivier, R. Brossier, S. Operto, and J. Virieux. Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated Newton approach. In *SEG Technical Program Expanded Abstracts*, pages 1198–1203, 2015.
- [15] L. Métivier, R. Brossier, S. Operto, and J. Virieux. Full waveform inversion and the truncated Newton method. *Siam Review*, 59(1):153–195, 2017. ISSN 0036-1445. 10.4208/cicp.OA-2018-008710.1137/16M1093239. URL <Go to ISI>://WOS:000396692500007.
- [16] S. Operto, Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Métivier, and J. Virieux. A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice. *The Leading Edge*, 32(9):1040–1054, 2013.
- [17] W. Y. Pan, K. A. Innanen, G. F. Margrave, M. C. Fehler, X. D. Fang, and J. X. Li. Estimation of elastic constants for HTI media using Gauss-Newton and full-Newton multiparameter full-waveform inversion. *Geophysics*, 81(5):R275–R291, 2016. ISSN 0016-8033. 10.4208/cicp.OA-2018-008710.1190/Geo2015-0594.1. URL <Go to ISI>://WOS:000392752200053.
- [18] W. Y. Pan, K. A. Innanen, and W. Y. Liao. Accelerating Hessian-free Gauss-Newton full-waveform inversion via l-BFGS preconditioned conjugate-gradient algorithm. *GEO-PHYSICS*, 82(2):R49–R64, 2017.
- [19] W. Y. Pan, Y. Geng, and K. A. Innanen. Interparameter trade-off quantification and reduction in isotropic-elastic full-waveform inversion: synthetic experiments and hussar land data set application. *Geophysical Journal International*, 213(2):1305–1333, 2018a. 10.4208/cicp.OA-2018-008710.1093/gji/ggy037. URL <http://dx.doi.org/10.1093/gji/ggy037>.
- [20] W. Y. Pan, K. A. Innanen, and Y. Geng. Elastic full-waveform inversion and parameterization analysis applied to walk-away vertical seismic profile data for unconventional

- (heavy oil) reservoir characterization. *Geophysical Journal International*, accepted, 2018b. 10.4208/cicp.OA-2018-008710.1093/gji/ggy087.
URL <http://dx.doi.org/10.1093/gji/ggy087>.
- [21] R. E. Plessix, P. Milcik, H. Rynja, A. Stopin, K. Matson, and S. Abri. Multiparameter full-waveform inversion: Marine and land examples. *The Leading Edge*, 32(9):1030–1038, 2013.
- [22] R. G. Pratt, C. Shin, and G. J. Hicks. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, 133(2):341–362, 1998. ISSN 0956-540X. URL <Go to ISI>://WOS:000073793700010.
- [23] M. S. Sambridge, A. Tarantola, and B. L. N. Kennett. An alternative strategy for nonlinear inversion of seismic wave-forms. *Geophysical Prospecting*, 39(6):723–736, 1991. ISSN 0016-8025. 10.4208/cicp.OA-2018-0087DOI 10.1111/j.1365-2478.1991.tb00341.x.
URL <Go to ISI>://WOS:A1991FZ72300001.
- [24] G. Shin, S. Jang, and D. J. Min. Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical Prospecting*, 49(5):592–606, 2001. ISSN 0016-8025. URL <Go to ISI>://WOS:000171314300009.
- [25] A. Tarantola. Inversion of seismic-reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266, 1984. ISSN 0016-8033. 10.4208/cicp.OA-2018-0087Doi 10.1190/1.1441754.
- [26] A. Tarantola. A strategy for nonlinear elastic inversion of seismic reflection data. *GEOPHYSICS*, 51(10):1893–1903, 1986. 10.4208/cicp.OA-2018-008710.1190/1.1442046.
URL <https://doi.org/10.1190/1.1442046>.
- [27] J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):Wcc1–Wcc26, 2009. ISSN 0016-8033. 10.4208/cicp.OA-2018-008710.1190/1.3238367.
- [28] R. Wu and K. Aki. Scattering characteristics of elastic waves by an elastic heterogeneity. *GEOPHYSICS*, 50(4):582–595, 1985. 10.4208/cicp.OA-2018-008710.1190/1.1441934.
URL <https://doi.org/10.1190/1.1441934>.
- [29] R. S. Wu and Y. C. Zheng. Non-linear partial derivative and its de wolf approximation for non-linear seismic inversion. *Geophysical Journal International*, 196(3):1827–1843, 2014. ISSN 0956-540X. 10.4208/cicp.OA-2018-008710.1093/gji/ggt496.
- [30] Z. D. Wu and T. Alkhalifah. Efficient scattering-angle enrichment for a nonlinear inversion of the background and perturbations components of a velocity model. *Geophysical Journal International*, online, 2017.
- [31] J. Z. Yang, Y. Z. Liu, and L. G. Dong. Simultaneous estimation of velocity and density in acoustic multiparameter full-waveform inversion using an improved scattering-integral approach. *Geophysics*, 81(6):R399–R415, 2016. ISSN 0016-8033. 10.4208/cicp.OA-2018-008710.1190/Geo2015-0707.1. URL <Go to ISI>://WOS:000392926400034.
- [32] Y. X. Yuan. Subspace methods for large scale nonlinear equations and nonlinear least squares. *Optimization and Engineering*, 10(2):207–218, 2009. ISSN 1389-4420. 10.4208/cicp.OA-2018-008710.1007/s11081-008-9064-0.
URL <Go to ISI>://WOS:000266497700005.
- [33] Y. X. Yuan and J. Stoer. A subspace study on conjugate-gradient algorithms. *Zeitschrift Fur Angewandte Mathematik Und Mechanik*, 75(1):69–77, 1995. ISSN 0044-2267. 10.4208/cicp.OA-2018-0087DOI 10.1002/zamm.19950750118. URL <Go to ISI>://WOS:A1995QD17600007.