Subspace method for multi-parameter FWI

Yu Geng, Kris Innanen and Wenyong Pan

ABSTRACT

Full waveform inversion aims to find the high-resolution subsurface model parameters. It is usually treaded as a nonlinear least square problem, and the minimum of the related misfit function is found by updating the model parameters. Simple gradient methods could mix different parameter types in the case of inversion with multi-parameter classes, which could lead to a poor convergence and strong dependence on the scaling of the different parameter types. Searching the step length in a subspace domain instead of treating the gradients of different parameters as the same could help solving this problem. The subspace used can be defined in a span of different sets of data or different parameter classes, which is a small amount of vectors compared to the whole model space. Using the subspace method, the basis vectors are needed to be defined first, and a local minimum is found in the spanned space to invert the perturbations. We are investigating this method to get a better update of density.

INTRODUCTION

In full waveform inversion (FWI) (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009), subsurface model parameters are found through the inversion scheme, which could be used to generate the predicted data closely resembling the recorded data under some physical assumption of the forward modelling, e.g., from acoustic case to viscoelastic anisotropic dynamics. Inversion of different classes of parameter (e.g., Fichtner, 2011; Operto et al., 2013; Plessix et al., 2013; Alkhalifah and Plessix, 2014; Pan et al., 2016), e.g., P-wave velocity, density, attenuation, shear-wave velocity and anisotropy, are required as a trend in the study of FWI, which leads to the development of multiparameter inversion. Similar to the monoparameter inversion under the acoustic assumption, which usually inverts only P-wave velocity, a misfit function is involved 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. Multiparameter inversion is much more complicated than monoparameter inversion, since the additional parameter classes increases the ill-posedness and the nonlinearity of the inverse problem. Different parameter classes can be more or less coupled, and it may be hard to distinguish the contribution of each parameter class to the change in the data. Mitigating the cross-talk between different parameter classes in the inversion becomes a very important topic in multiparameter inversion. Studies have shown that the Hessian operator usually contains some information of the coupling between different parameter classes. Different ways of cooperating the inverse of Hessian operator, especially in the multiparameter cases, are proposed to better uncouple different parameter classes in the inversion, such as precondition the gradient using pseudo Hessian matrix, quasi-Newton method, truncated Newton method and so on. Hierarchical strategies can be applied to successively invert different parameter classes to mitigate the ill-posedness of FWI (Jeong et al., 2012), which will require more computation.

In both gradient-based methods and Newton-type methods (e.g., Virieux and Operto, 2009; Pratt et al., 1998; Métivier et al., 2013, 2014; Yang et al., 2016; Pan et al., 2017), line searching is usually necessary to scale the descent direction so that the method can be globally convergent. One scalar is the found for all parameter classes regardless their own contributions to the data. Distinguish the contribution of each parameter class during the updating could be helpful in multiparameter inversion. Application of subspace method in large inverse problems was first discussed by (Kennett et al., 1988; Sambridge et al., 1991), to adjust the descent direction according to different parameter classes' contribution. Baumstein (2014) show that using an extended subspace method in multiparameter inversion can help to mitigate the cross-talk as well. In subspace FWI, the 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 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 compared to Newton-type methods. In this study, we evaluate different basis vectors, constructed from the gradient of different parameter classes, and related Hessian-vectors, to construct a better multiparameter inversion. Although we use acoustic wave equation with varying density for the forward modelling and invert velocity and density, the application to inversion of other parameterizations and elastic wave equation can be easily extended.

REVIEW OF SUBSPACE METHOD

In this study, we will use the frequency-space domain acoustic wave equation to describe the wave motion,

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

where ρ is the density and v is the velocity. Write the model parameters with different types into one vector $\mathbf{m}$, discretized wave equation can be written in matrix form as

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

The forward modeling (2) describes a nonlinear relationship between the wavefield and the model. The inversion problem can be seen as an optimization problem, which is to find a model $\mathbf{m}$ to minimize the misfit functional $\phi(\mathbf{m})$

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

where $\mathbf{d}_{obs}(\mathbf{x}_s,\omega)$ is the observed data for each source location $\mathbf{x}$ for one frequency ω, and $\mathbf{d}_{syn}(\mathbf{m},\mathbf{x}_s,\omega)=\mathbf{R}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$ is the synthetic data generated using the forward modeling (2) in the current model $\mathbf{m}$ and sampled with an operator $\mathbf{R}$ on the receiver locations. $\delta\mathbf{d}(\mathbf{m},\mathbf{x}_s,\omega)$ is the data residual, which is defined as the difference between the observed data and the synthetic data.

Expanding the misfit functional (3) up to second order around the vicinity of the model $\mathbf{m}$,

$$\phi(\mathbf{m}+\delta\mathbf{m})=\phi(\mathbf{m})+<\mathbf{g},\delta\mathbf{m}>+\frac{1}{2}<\mathbf{H}\delta\mathbf{m},\delta\mathbf{m}>+O\left(\|\delta\mathbf{m}\|^3\right), \tag{4}$$

where $\mathbf{g}$ and $\mathbf{H}$ are the gradient and the Hessian operator of the misfit function, respectively. Since the misfit function is usually non-quadratic, based on a local quadratic approximation as shown in (4), the model can then be updated iteratively by a perturbation as

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \alpha_n \delta\mathbf{m}. \tag{5}$$

The perturbation can be determined by a descent direction, which can be the opposite of the gradient in the gradient-based methods, or it can be the solution of the perturbation from the linearized inversion in the Newton-type method

$$\begin{aligned} \delta\mathbf{m} &= -\mathbf{g}, \quad \text{and} \\ \delta\mathbf{m} &= -\mathbf{H}^{-1}\mathbf{g}. \end{aligned} \tag{6}$$

However, all different parameters are updated with the related descent direction scaled by a step-length α as in (5). This can be treated 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 down the descent direction. The step-length is a constant for each parameter type, and the updates of each parameter is governed by the properties of the overall descent direction at each iteration, instead of its own direction.

In the case of single parameter, subspace methods can also be used. Usually gradient of current iteration and previous step information are combined to construct a subspace, where a step length is found in a subspace of dimension M, which is small. However, it is usually a trade-off between the increasing in computational cost per iteration and the possible decrease in number of iteration. Under this definition, conjugate gradient method can be seen as a subspace method, which uses previous step and current gradient to construct a 2D subspace optimization, so as the limited memory quasi-Newton method. In these methods, the search directions are spanned in a lower dimensional space, which is at least 2, compared to the steepest descent method, which searches a 1D step-length in the full space (e.g., Yuan and Stoer, 1995; Yuan, 2009).

In the case of multi-parameter, either different datasets can be used to construct a subspace, model space can also be divided into subspace. In this study, we consider the partition the gradient into contributions of each parameter to distinguish the contribution of each gradient. We will study the inversion of acoustic case to invert both velocity and density (different parameterization can be applied also) in frequency domain. Suppose that the perturbation can be written as a combination in a space spanned by n basis vectors $\{\mathbf{a}_j\}$

$$\delta\mathbf{m} = \sum_{j=1}^{n} \mu_j \mathbf{a}_j. \tag{7}$$

Insert these perturbations (7) into the expansion of the misfit functional (4)

$$\phi(\mathbf{m} + \delta\mathbf{m}) = \phi(\mathbf{m}) + \sum_{j=1}^{n} \mu_j < \mathbf{g}, \mathbf{a}_j > + \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} \mu_j \mu_i < \mathbf{H}\mathbf{a}_j, \mathbf{a}_i > + \dots . \tag{8}$$

The coefficients can be found using

$$\mu_j < \mathbf{g}, \mathbf{a}_j > + \sum_{j=1}^{n} \mu_j < \mathbf{H}\mathbf{a}_j, \mathbf{a}_i > \approx 0. \tag{9}$$

Written in matrix form

$$\mathbf{A}^T\mathbf{g} + \mathbf{A}^T\mathbf{H}\mathbf{A}\mu \approx 0. \tag{10}$$

The coefficients can then be determined from the projection of the gradient and the Hessian onto the subspace in the form

$$\mu = -\left(\mathbf{A}^T\mathbf{H}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{g}. \tag{11}$$

Since the subspace is only n dimension, $\mathbf{A}^T\mathbf{H}\mathbf{A}$ is a $n \times n$ matrix and simple to invert. When the second derivative term in the Hessian can be neglected, the approximate Hessian can be used in the equation to evaluate the coefficient.

CHOICE OF SUBSPACE BASIS VECTORS

subspace basis vectors for linear updates

2D choice

It is straight forward to choose the descent direction (opposite of gradient) of the each parameter as the basis vectors, in which case, the descent direction of each parameter needs to be extended into the whole model space, e.g., in the case of two parameters v and ρ , the basis vectors are the extension the gradient of each parameter in the whole model space,

$$\mathbf{a}_1 = \begin{bmatrix} -\mathbf{g}_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ -\mathbf{g}_\rho \end{bmatrix}. \tag{12}$$

The Hessian matrix can be written as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{vv} & \mathbf{H}_{v\rho} \\ \mathbf{H}_{\rho v} & \mathbf{H}_{\rho\rho} \end{bmatrix} \tag{13}$$

The coefficients are

$$\mu = \begin{bmatrix} \mathbf{g}_v^T\mathbf{H}_{vv}\mathbf{g}_v & \mathbf{g}_v^T\mathbf{H}_{v\rho}\mathbf{g}_\rho \\ \mathbf{g}_\rho^T\mathbf{H}_{\rho v}\mathbf{g}_v & \mathbf{g}_\rho^T\mathbf{H}_{\rho\rho}\mathbf{g}_\rho \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}_v^T\mathbf{g}_v \\ \mathbf{g}_\rho^T\mathbf{g}_\rho \end{bmatrix} \tag{14}$$

which is easy to be calculated. The calculation is involving a calculation of a Hessian-vector as in the Hessian-free Newton method. In this case, two Hessian-vectors are needed for each calculation of the coefficient, and a 2×2 matrix is inverted. Compared to the steepest descent method, which usually involving a line search scheme, and in a conjugate gradient method for linear problem, e.g., with in the inner loop of the truncated Gauss-Newton method, and the step-length is calculated

$$\mu = \frac{< \mathbf{g}, \mathbf{g} >}{< \mathbf{H}\mathbf{g}, \mathbf{g} >}, \tag{15}$$

The step length for update each parameter type is different, and the model is updated as

$$\delta v = \mu_v \mathbf{g}_v, \quad \delta\rho = -\mu_\rho \mathbf{g}_\rho. \tag{16}$$

4D choice

These basis vectors can also be extended to a way to pretend the possible leakage between gradient components, e.g.,

$$\mathbf{a}_1 = \begin{bmatrix} -\mathbf{g}_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} -\mathbf{g}_\rho \\ 0 \end{bmatrix} . \mathbf{a}_3 = \begin{bmatrix} 0 \\ -\mathbf{g}_v \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 \\ -\mathbf{g}_\rho \end{bmatrix}. \tag{17}$$

In this case,

$$\mu = \begin{bmatrix} \mathbf{g}_v^T\mathbf{H}_{vv}\mathbf{g}_v & \mathbf{g}_v^T\mathbf{H}_{vv}\mathbf{g}_\rho & \mathbf{g}_v^T\mathbf{H}_{v\rho}\mathbf{g}_v & \mathbf{g}_v^T\mathbf{H}_{v\rho}\mathbf{g}_\rho \\ \mathbf{g}_\rho^T\mathbf{H}_{vv}\mathbf{g}_v & \mathbf{g}_\rho^T\mathbf{H}_{vv}\mathbf{g}_\rho & \mathbf{g}_\rho^T\mathbf{H}_{v\rho}\mathbf{g}_v & \mathbf{g}_\rho^T\mathbf{H}_{v\rho}\mathbf{g}_\rho \\ \mathbf{g}_v^T\mathbf{H}_{\rho v}\mathbf{g}_v & \mathbf{g}_v^T\mathbf{H}_{\rho v}\mathbf{g}_\rho & \mathbf{g}_v^T\mathbf{H}_{\rho\rho}\mathbf{g}_v & \mathbf{g}_v^T\mathbf{H}_{\rho\rho}\mathbf{g}_\rho \\ \mathbf{g}_\rho^T\mathbf{H}_{\rho v}\mathbf{g}_v & \mathbf{g}_\rho^T\mathbf{H}_{\rho v}\mathbf{g}_\rho & \mathbf{g}_\rho^T\mathbf{H}_{\rho\rho}\mathbf{g}_v & \mathbf{g}_\rho^T\mathbf{H}_{\rho\rho}\mathbf{g}_\rho \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}_v^T\mathbf{g}_v \\ \mathbf{g}_\rho^T\mathbf{g}_v \\ \mathbf{g}_v^T\mathbf{g}_\rho \\ \mathbf{g}_\rho^T\mathbf{g}_\rho \end{bmatrix} \tag{18}$$

In this case, four Hessian-vectors are needed to be calculated and an 4×4 matrix is inverted to calculate the coefficients for four basis vectors.

6D choice

The rate of change of the ascent vectors can also be used to construct the basis vectors, e.g.,

$$\begin{aligned} \mathbf{a}_1 &= \begin{bmatrix} -\mathbf{g}_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ -\mathbf{g}_\rho \end{bmatrix}, \\ \mathbf{a}_3 &= \begin{bmatrix} \mathbf{H}_{vv} & \mathbf{H}_{v\rho} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{g}_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 & 0 \\ \mathbf{H}_{\rho v} & \mathbf{H}_{\rho\rho} \end{bmatrix} \begin{bmatrix} \mathbf{g}_v \\ 0 \end{bmatrix}, \\ \mathbf{a}_5 &= \begin{bmatrix} \mathbf{H}_{vv} & \mathbf{H}_{v\rho} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{g}_\rho \end{bmatrix}, \quad \mathbf{a}_6 = \begin{bmatrix} 0 & 0 \\ \mathbf{H}_{\rho v} & \mathbf{H}_{\rho\rho} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{g}_\rho \end{bmatrix}. \end{aligned} \tag{19}$$

As pointed out in Baumstein's study, when the misfit function is not locally quadratic, curvature information obtained from the Hessian may be far away from the global minimum, and the coefficients obtained may not lead to an improved search direction. Also, when using the Hessian in the constructing of the basis vectors, computation cost may be much higher, since more Hessian-vectors are needed.

Subspace basis vectors for nonlinear update

When considering the second-order scattering in the Hessian operator as correction in the gradient, in the approximate Newton method, the higher-order perturbations can be used as the basis vectors beside the gradient for each parameter type. In this case, the Hessian can be written as summation of two parts

$$\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2, \tag{20}$$

where $\mathbf{H}_1$ contains the second-order partial derivative of the data 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 (5) to

$$\delta\mathbf{m} = -\mathbf{H}_2^{-1}\left(\mathbf{g} - \mathbf{H}_1\mathbf{H}_2^{-1}\mathbf{g}\right). \tag{21}$$

To calculate this model perturbation, the inverse of the Gauss-Newton Hessian $\mathbf{H}_2^{-1}$ is needed, to avoid the huge cost of calculation of this term directly, quasi-Newton method and 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., truncated Gauss-Newton method. The basis vectors can be a combination between the gradient vectors for each parameter class and also its related nonlinear perturbations,

$$\mathbf{a}_1 = \begin{bmatrix} -\mathbf{g}_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} (\mathbf{H}_1\delta\mathbf{m}_1)_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 \\ -\mathbf{g}_\rho \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 \\ (\mathbf{H}_1\delta\mathbf{m}_1)_\rho \end{bmatrix}. \tag{22}$$

Compared to the 4D subspace constructed from the gradient vectors (19) one more Hessian-vector $\mathbf{H}_1\delta\mathbf{m}_1$ is needed during the calculation of the coefficients.

To find the update using subspace basis (22), a perturbation model $\delta\mathbf{m}_1$ is needed to form the basis vectors $\mathbf{a}_2$ and $\mathbf{a}_4$. The calculation of this perturbation model could be essential to the whole nonlinear update, since cross-talk artefacts should be removed by the inverse of Gauss-Newton Hessian operator to avoid introducing new artefacts. Moreover, the effect of inverse of the Gauss-Newton Hessian operator may not be ignored since cross-talk artefacts still exist in the model updates founded within subspace basis vectors (22). Therefore, directly using perturbation model $\delta\mathbf{m}_1$ to construct the basis vectors could be a better choice

$$\begin{aligned} \mathbf{a}_1 &= \begin{bmatrix} (\delta\mathbf{m}_1)_v \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} -\mathbf{H}_2^{-1}(\mathbf{H}_1\delta\mathbf{m}_1)_v \\ 0 \end{bmatrix}, \\ \mathbf{a}_3 &= \begin{bmatrix} 0 \\ (\delta\mathbf{m}_1)_\rho \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 \\ -\mathbf{H}_2^{-1}(\mathbf{H}_1\delta\mathbf{m}_1)_\rho \end{bmatrix}. \end{aligned} \tag{23}$$

In this study, we use l-BFGS to approximate the inverse of the Gauss-Newton Hessian as a precondition for the descent direction, and pseudo Hessian can also be used with the assume that for the multiparameter case, each block in the Hessian for different parameter classes is diagonal (Shin et al., 2001; Innanen, 2014; Métivier et al., 2015). In the cases where the basis vectors may have cross-coupling with each other, e.g., using gradients of different parameter classes (17), or using the Hessian to construct the basis vectors for each parameter classes as in (19) and (22), orthogonalizing the basis vectors may be needed to avoid linear dependence between the different basis vectors related with one parameter class

$$\hat{\mathbf{a}}_2 = \mathbf{a}_2 - \frac{\mathbf{a}_1\mathbf{a}_2}{\mathbf{a}_1\mathbf{a}_1}\mathbf{a}_1. \tag{24}$$

APPLICATION OF SUBSPACE METHOD IN FWI

Using subspace methods in FWI instead of simple line searching method can help altering the descent direction by emphasizing the contribution of each parameter classes. Compared to the gradient based method, the Hessian operator is involved in the calculation of subspace vector coefficients. Compared to Newton-type method, subspace method may be able to better cooperate the contribution of the multi-scattering related model perturbation into the descent direction for each parameter classes. Therefore, the usage of subspace method in multiparameter FWI may help to mitigate the cross-talk between different parameter classes, since it can update different parameter classes with different step length, and it can include the multi-scattering related perturbations as a correction to the single-scattering related perturbations for different parameter classes.

Advantage of the subspace method

1. Gradients with different parameters are considered individually. A least-square inversion is performed within the subspace, which is constructed from the dependence of all different parameters.

2. Coefficients are calculated using the Hessian-vector, which is much cheaper than Newton/truncated Newton/Gauss-Newton method, but could be compared with truncated Gauss-Newton/Newton method.

3. May be useful for attenuation of crosstalk in multi-parameter FWI.

Disadvantage of the subspace method

1. The theory relies on the assumption that the Hessian correctly captures the behavior of the misfit function. When the misfit function is not locally quadratic, it is not guaranteed that the global minimum of the misfit can be found.

2. Approximate Hessian is usually used instead of the exact Hessian, which may make the Hessian itself inaccurate.

3. It is possible that the model perturbation calculated from (11) is too large that the misfit function is not quadratic anymore. In this case, another scaling factor may be needed to modify the coefficients so that the updated model can still help the misfit function to converge to the quadratic minimum.

EXAMPLES AND PRACTICAL ISSUES

We will test the application of subspace method in both gradient-type and Newton-type FWI to invert velocity and density simultaneously in both transmission and reflection cases. Steep descent method and truncated Newton method are used, where the line searching are replaced with the subspace method as proposed in this paper.

Transmission case: Two Gaussian balls

Using subspace method instead of line search

Figure 1 shows the true and initial models for both velocity and density. The sources are at the top of the model, and the receivers are at the bottom of the model. 5 outer iterations are used for both steep descent method and truncated Gauss-Newton method, where 10 CG iterations are used to solve the linearized inversion problem. Three data set for the frequencies 8, 10 and 15 Hz are used for the inversion. We test both 2D and 4D subspace method, and the inversion results are shown as in Figure 2. Gauss-Newton Hessian is used. Figure 3 shows the profile of both velocity and density along $z = 0.25$ km. To compare the convergence of the subspace FWI with the original FWI with line searching, we show both the misfit and model errors vs. iterations as in Figure 4.

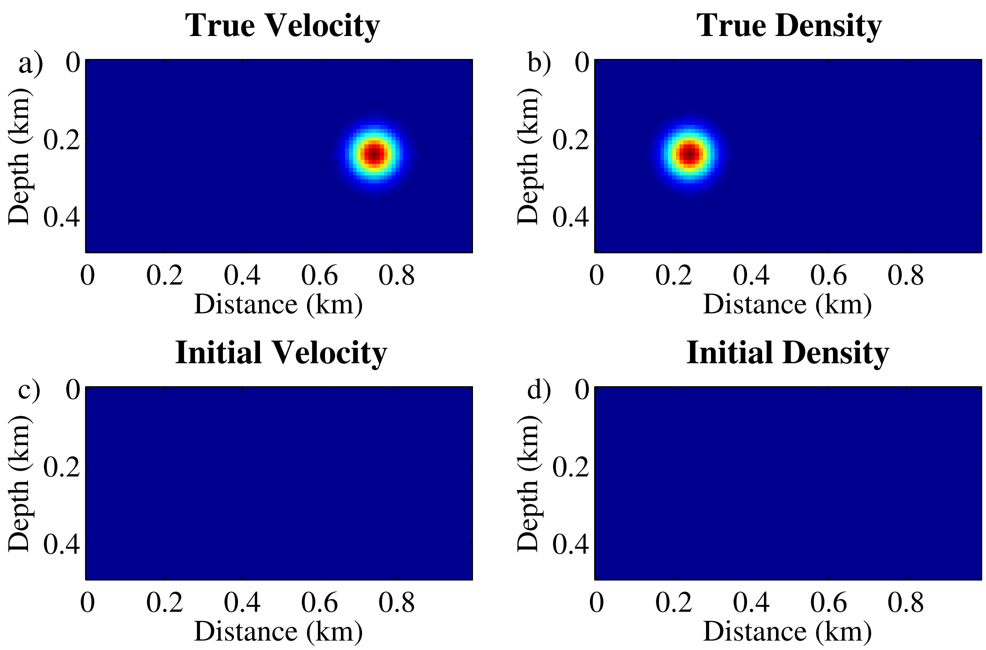
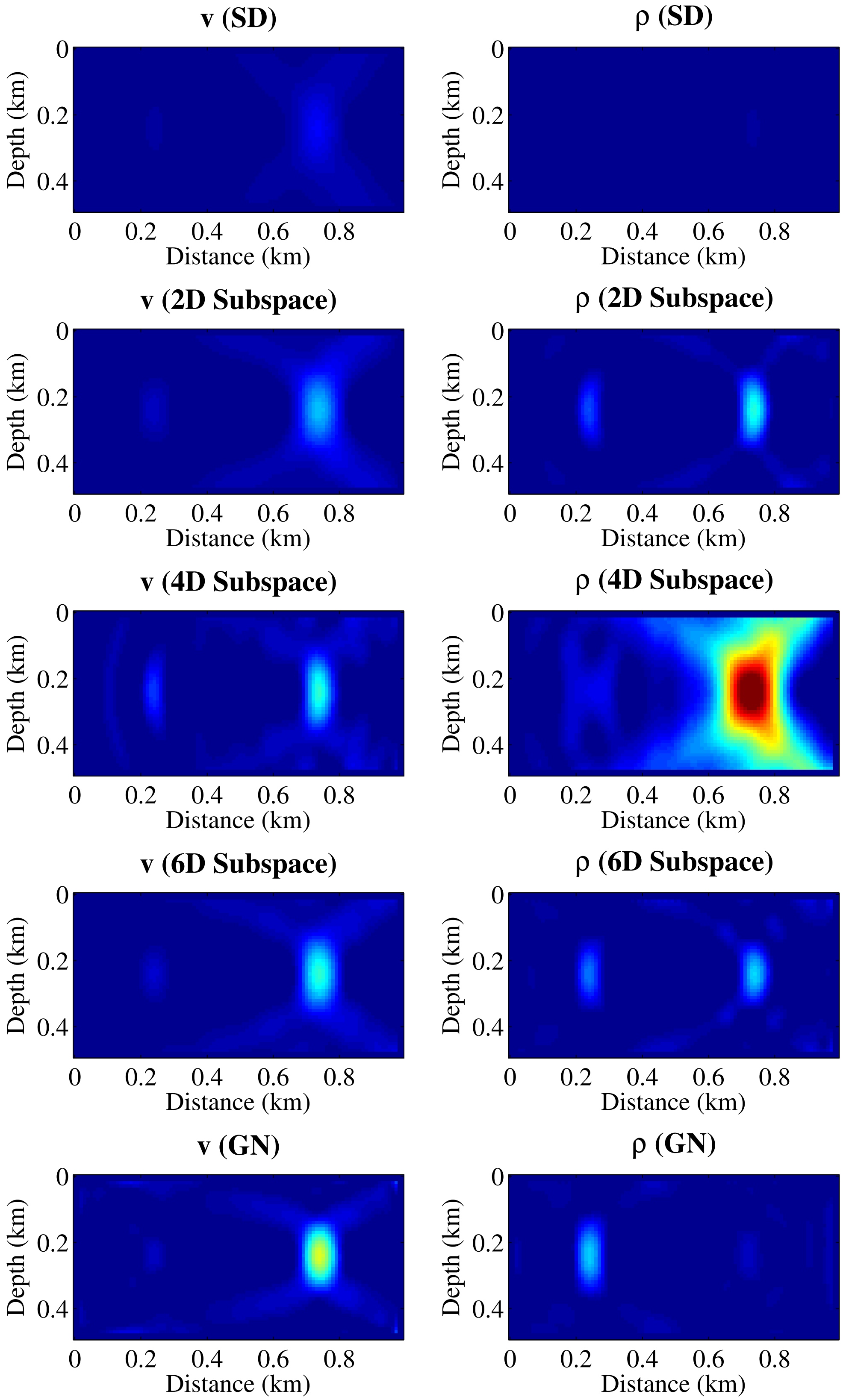


FIG. 1. True velocity a) and density b). Initial velocity c) and density d).

From the inversion results, we can see that, applying subspace method to the steepest descent method can provide a better update for the velocity with both 2D and 4D gradient based basis vectors, however, the updating for density is getting worse due to the cross-talk between velocity and density. It can be explained since the radiation pattern for velocity is isotropic, but for density the radiation pattern is strong at small scattering angle, which makes it hard to update density in the transmission case. Besides, recorded data is more sensitive to velocity, and barely response to the density change, which could results a better updates for velocity, but not for density. Cross talk between velocity and density is strong and positively related to each other, so compared to 2D subspace, using the 4D subspace method won't provide much help to better update any parameter, and it could slow down the convergence in certain point. The steepest descent subspace method can provide a better convergence only for the first iteration when compared to the Gauss-Newton method, instead, when combining subspace method with Gauss-Newton method, the cross-talk between velocity and density is reduced, and both 2D and 4D subspace method can provide a


v (SD)
ρ (SD)
v (2D Subspace)
ρ (2D Subspace)
v (4D Subspace)
ρ (4D Subspace)
v (6D Subspace)
ρ (6D Subspace)
v (GN)
ρ (GN)
Depth (km)
Distance (km)
0
0.2
0.4
0.6
0.8

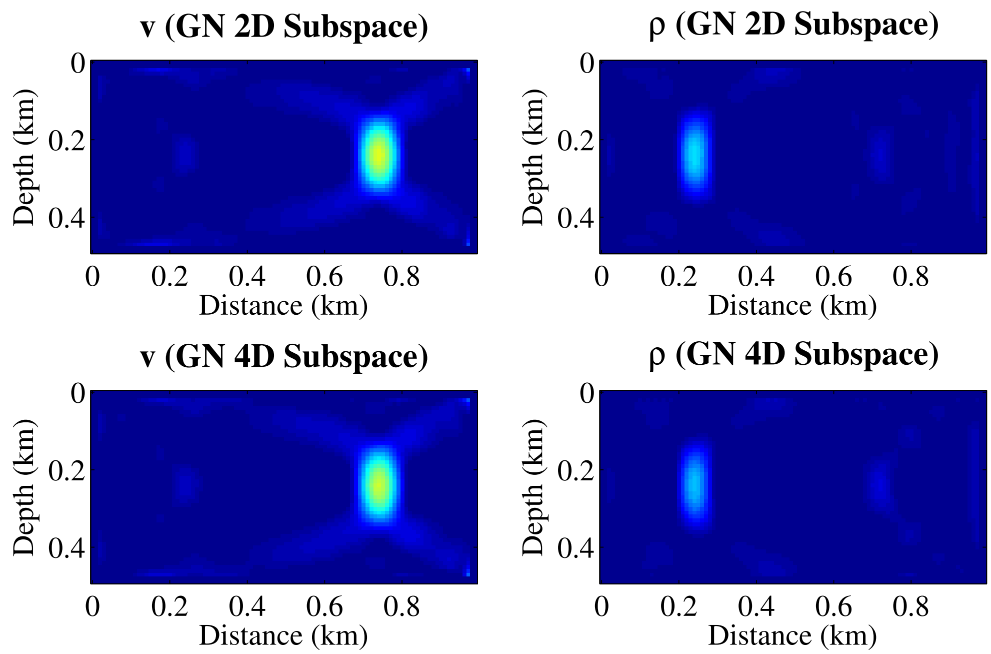


FIG. 2. Inverted velocity and density using different methods.

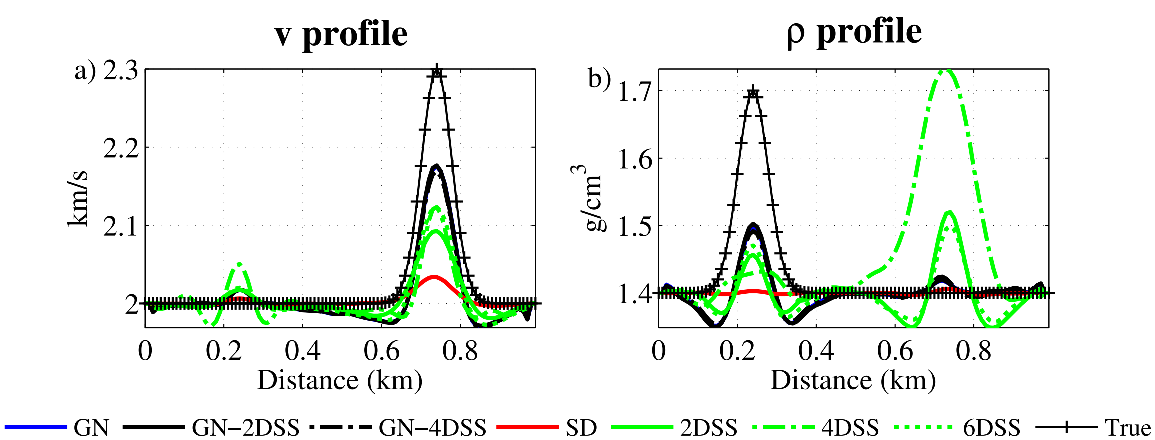


FIG. 3. Profile for velocity and density at $z = 0.25$ km.

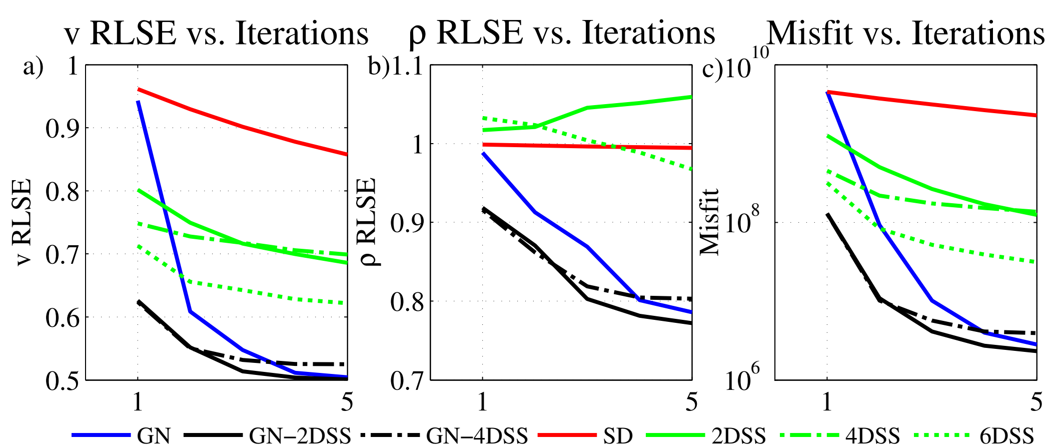


FIG. 4. Comparison of the convergence of the truncated Newton method with/without subspace method.

better inversion results, especially for density.

Nonlinear update

We then use nonlinear term to construct the basis vectors. Maximum 10 inner iterations and 20 outter iterations are used for the inversion. Figure 5 compare the results of different methods. Figure 6 show the related profiles of velocity and density at $z = 0.25$ km. It can be observed that, 1), using full Hessian instead of the Gauss-Newton Hessian in the truncated Newton method doesn't provide a better update, epsecially for velocity; 2), using model perturabtion obtained from a Gauss-Newton update and its related nonlinear term with or without l-BFGS pre-conditioning, the inversion results converge faster in the early stage for both velocity and density; 3), using gradient and the nonlinear term (model perturbation approximated with l-BFGS update) to construct the subspace basis vector, the inversion of velocity is better than the truncated Newton method, however, due to the assumption that the inverse of Gauss-Newton Hessian is identity matrix, the cross-talk artifacts make the inversion of density worse and fail to converge to the true model.

Reflection case: Marmousi

It is more complex to apply the subspace method in the reflection case. We take only a small area of the original Marmousi-II model as our true model, and the initial model is obtained by smoothing the true model with a Gaussian smoothing window under the water layer, as shown in Figure 8. The new model is in the size of 81×161 grid nodes, with grid intervals of 20 m at each direction. 4 frequencies $(3,\ 5,\ 8,\ 12$ Hz) are inverted simultaneously, and a maximum of 20 iterations are performed for all the methods, where 10 inner iterations are used for the truncated Gauss-Newton type methods. Inversion results using different methods (2D subspace method, 6D subspace method, Gauss-Newton method, and Gauss-Newton method combined with 2D subspace method) are shown in Figure 9, and Figure 10 shows the convergence profiles of these four methods. It can be seen that using subspace method can improve the updating of the velocity, but taking the price of a overestimating of the density.

CONCLUSIONS

In this study, we are interested in the application of the subspace method on simultaneously updating the velocity and density. Subspace method can be used instead of line searching in the traditional implementation of FWI to obtain step lengths for different parameter class. Gauss-Newton Hessian product with a vector is involved to find the local minimum in the spanned space. We studied different basis vectors to construct the spanned space, with nonlinear perturbations obtained from higher-order scattering involved. The behavior of the subspace methods for both linear updates and nonlinear updates are compared with traditional FWI methods. The subspace methods have better convergence rate, as well as better reconstruction of the velocity model. The reconstruction of density model, however, could still be effected by the cross-talk artifacts, when Hessian is not considered in the inversion.

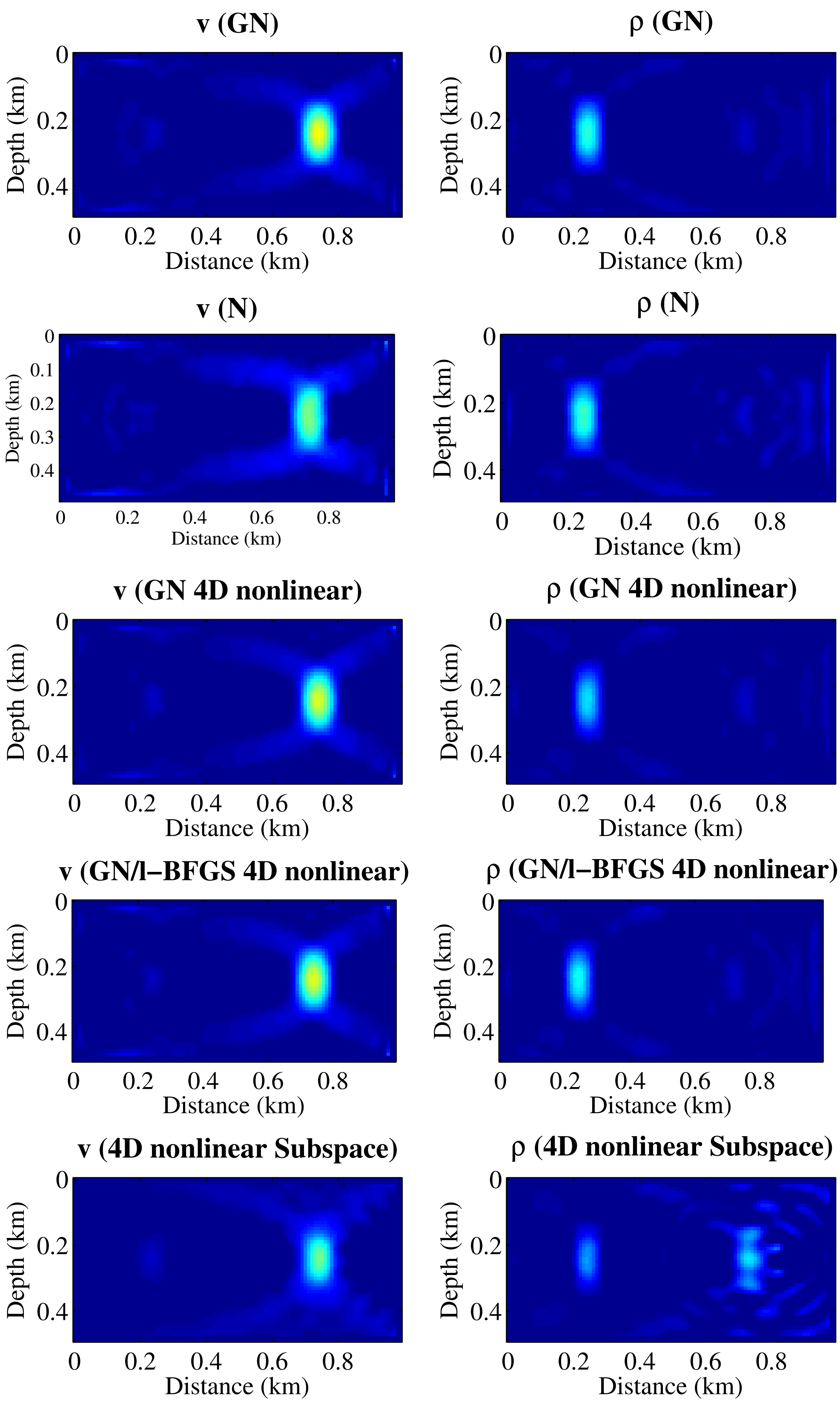


FIG. 5. Inverted velocity and density using different methods.

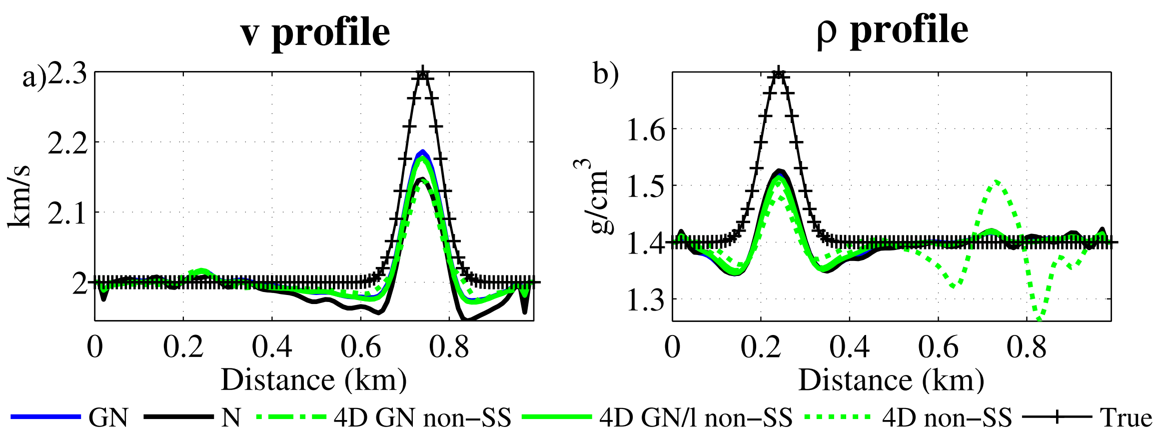


FIG. 6. Profile for velocity and density at $z = 0.25$ km.

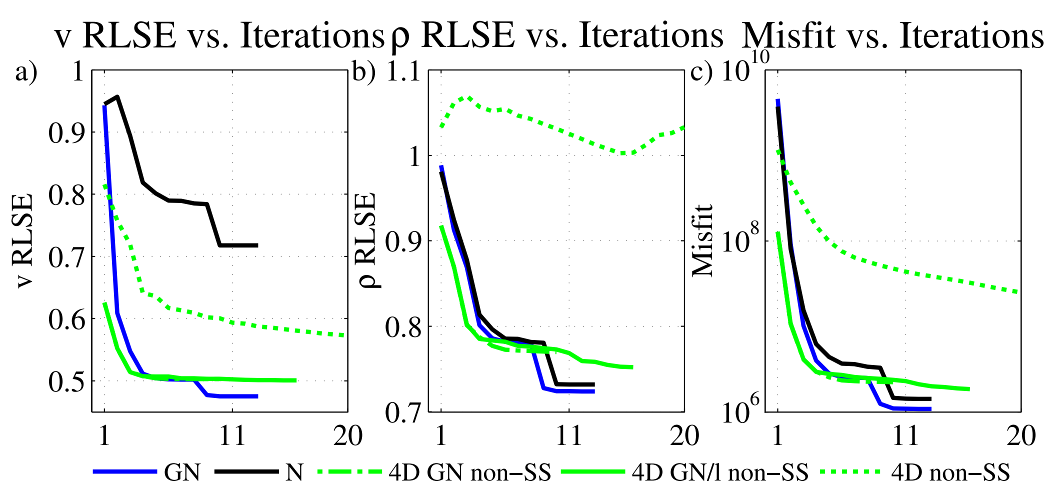


FIG. 7. Comparison of the convergence of the truncated Newton method and nonlinear subspace method.

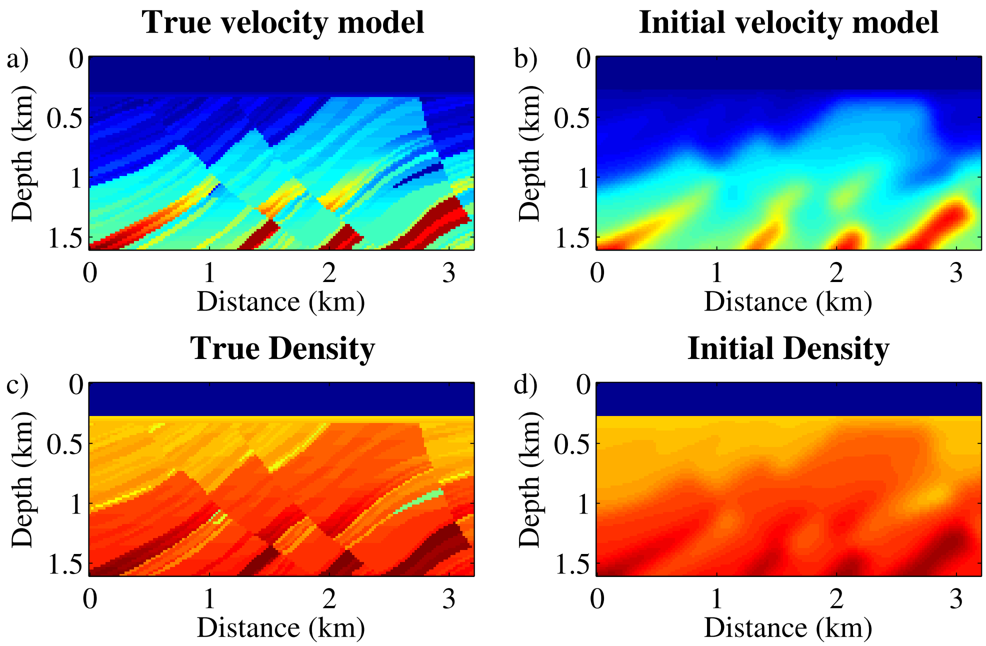


FIG. 8. True velocity a) and density b). Initial velocity c) and density d).

ACKNOWLEDGMENT

We thank the sponsors of CREWES for support. This work was funded by CREWES and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 379744-08.

REFERENCES

Alkhalifah, T., and Plessix, R. E., 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study: Geophysics, **79**, No. 3, R91–R101.
URL `<Go to ISI>://WOS:000338322900029`

Baumstein, A., 2014, Extended subspace method for attenuation of crosstalk in multi-parameter full wavefield inversion, *in* SEG Technical Program Expanded Abstracts, 1121–1125.

Fichtner, A., 2011, Full Seismic Waveform Modelling and Inversion: Springer-Verlag Berlin Heidelberg.

Innanen, K. A., 2014, Seismic avo and the inverse hessian in precritical reflection full waveform inversion: Geophysical Journal International, **199**, No. 2, 717–734.
URL `<Go to ISI>://WOS:000343411700008`

Jeong, W., Lee, H. Y., and Min, D. J., 2012, Full waveform inversion strategy for density in the frequency domain: Geophysical Journal International, **188**, No. 3, 1221–1242.
URL `<Go to ISI>://WOS:000300567400033`

Kennett, B. L. N., Sambridge, M. S., and Williamson, P. R., 1988, Subspace methods for large inverse problems with multiple parameter classes: Geophysical Journal-Oxford, **94**, No. 2, 237–247.
URL `<Go to ISI>://WOS:A1988P610200005`

Lailly, P., 1983, 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, 206–220.

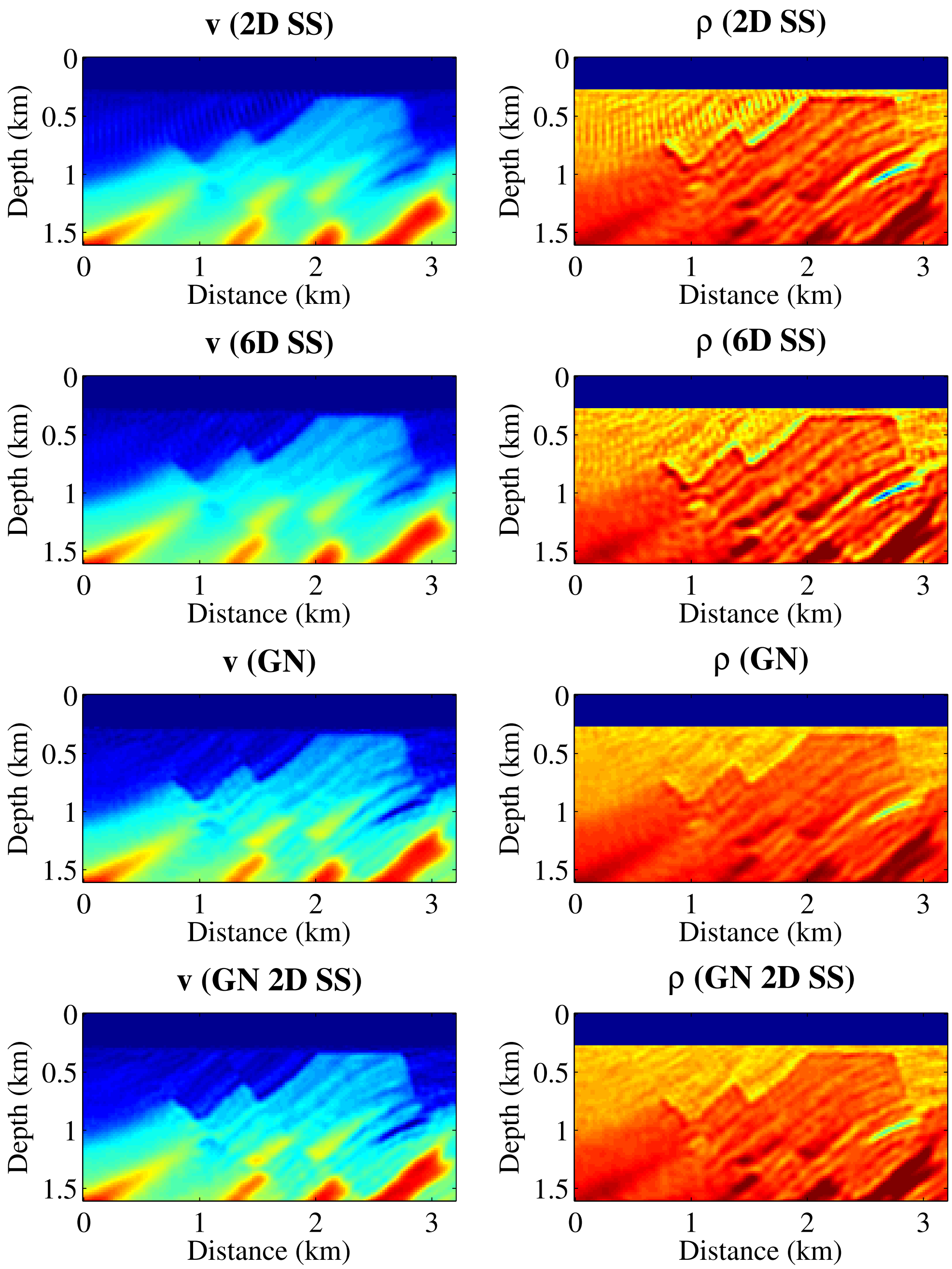


FIG. 9. Inverted velocity and density using different methods. Results are plotted in the same scale of the true model.

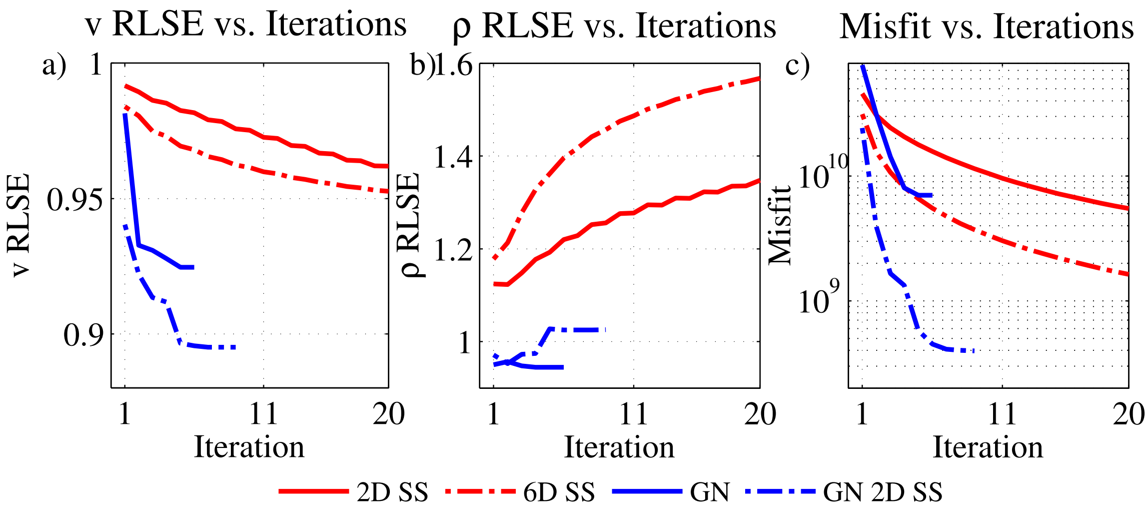


FIG. 10. Comparison of the convergence of the truncated Newton method and nonlinear subspace method.

Métivier, L., Bretaudeau, F., Brossier, R., Operto, S., and Virieux, J., 2014, Full waveform inversion and the truncated newton method: quantitative imaging of complex subsurface structures: Geophysical Prospecting, **62**, No. 6, 1353–1375.
URL `<Go to ISI>://WOS:000344222300013`

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, *in* SEG Technical Program Expanded Abstracts, 1198–1203.

Métivier, L., Brossier, R., Virieux, J., and Operto, S., 2013, Full waveform inversion and the truncated newton method: Siam Journal on Scientific Computing, **35**, No. 2, B401–B437.
URL `<Go to ISI>://WOS:000318404100035`

Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Métivier, L., and Virieux, J., 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: The Leading Edge, **32**, No. 9, 1040–1054.

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

Pan, W. Y., Innanen, K. A., Margrave, G. F., Fehler, M. C., Fang, X. D., and Li, J. X., 2016, Estimation of elastic constants for hti media using gauss-newton and full-newton multiparameter full-waveform inversion: Geophysics, **81**, No. 5, R275–R291.
URL `<Go to ISI>://WOS:000392752200053`

Plessix, R. E., Milcik, P., Rynja, H., Stopin, A., Matson, K., and Abri, S., 2013, Multiparameter full-waveform inversion: Marine and land examples: The Leading Edge, **32**, No. 9, 1030–1038.

Pratt, R. G., Shin, C., and Hicks, G. J., 1998, Gauss-newton and full newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, No. 2, 341–362.
URL `<Go to ISI>://WOS:000073793700010`

Sambridge, M. S., Tarantola, A., and Kennett, B. L. N., 1991, An alternative strategy for nonlinear inversion of seismic wave-forms: Geophysical Prospecting, **39**, No. 6, 723–736.
URL `<Go to ISI>://WOS:A1991FZ72300001`

Shin, G., Jang, S., and Min, D. J., 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, No. 5, 592–606.
URL `<Go to ISI>://WOS:000171314300009`

Tarantola, A., 1984, Inversion of seismic-reflection data in the acoustic approximation: Geophysics, **49**, No. 8, 1259–1266.

Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, No. 6, Wcc1–Wcc26.

Yang, J. Z., Liu, Y. Z., and Dong, L. G., 2016, Simultaneous estimation of velocity and density in acoustic multiparameter full-waveform inversion using an improved scattering-integral approach: Geophysics, **81**, No. 6, R399–R415.
URL `<Go to ISI>://WOS:000392926400034`

Yuan, Y. X., 2009, Subspace methods for large scale nonlinear equations and nonlinear least squares: Optimization and Engineering, **10**, No. 2, 207–218.
URL `<Go to ISI>://WOS:000266497700005`

Yuan, Y. X., and Stoer, J., 1995, A subspace study on conjugate-gradient algorithms: Zeitschrift Fur Angewandte Mathematik Und Mechanik, **75**, No. 1, 69–77.
URL `<Go to ISI>://WOS:A1995QD17600007`