Preconditioned elastic full waveform inversion using approximated Hessian matrix

Ettore Biondi, Guillaume Barnier, and Biondo Biondi

ABSTRACT

We describe a simple way of estimating the elements of the Hessian matrix for elastic full waveform inversion (FWI) problems by applying it to spikes in the model space. We explain how to use this estimated matrix to precondition an elastic isotropic FWI problem and show results from a simple one layer model and a more complex subsurface elastic earth. We observe that already a main diagonal approximation is able to increase the inverse problem's convergence rate. This gradient preconditioning allows us to obtain a meaningful inversion result for the deeper part of the model in fewer iterations.

INTRODUCTION

Different preconditioning methods for FWI have been proposed in recent years. For example, Alkhalifah (2015b) describes a gradient preconditioning algorithm to avoid cycle skipping during FWI optimization. Based on the analysis performed by ten Kroode (2012) for Kirchhoff migration, Huang et al. (2016) shows how filtering FWI gradients in extended-image space can mimic the effect of a Newton optimization step. These two methods have been shown to be successful in the case of single-parameter inversions or when mostly the traveltimes are affected by the presence of multiple parameters (Alkhalifah, 2015a). However, it is missing a study of their advantage when also seismic wave amplitudes are influenced by different optimization parameters (e.g., elastic FWI).

Given the current interest in solving more complex and computationally intense multi-parameter FWI problems, such as fully orthorhombic elastic waveform inversion optimization (Albertin et al., 2016), it is critical to find methods to improve the convergence rate of any inversion algorithm employed. For this reason many authors have recognized the fundamental role of Hessian matrices as preconditioner for multi-parameter FWI problems (Forgues and Lambaré, 1997; Operto et al., 2013; Innanen, 2014). Tang and Lee (2015) show a simple way of estimating the Hessian matrix and demonstrate its value as preconditioner in a vertical transverse isotropic FWI example. In their implementation they take advantage of the sparse structure of the Gauss-Newton component of this matrix and compute its elements through matrix applications to spikes contained in the model space.
Korta et al. (2013) show how a block-diagonal Hessian approximation can improve

convergence rate during FWI optimization when compressional and shear wave velocities are inverted simultaneously. Similarly, using a phase-encoding approach to estimate the matrix elements (Tang, 2008), Deuzeman and Plessix (2015) demonstrate how a similar Hessian approximation speed-ups the optimization algorithm on simple subsurface models.

In this study we simultaneously invert for elastic wave velocities and density following the same approach proposed by Tang and Lee (2015) to estimate the Gauss-Newton Hessian elements. We show elastic FWI results on a complex subsurface elastic model and compare a main-diagonal approximation to un-preconditioned optimization.

THEORY

The 2D velocity-stress formulation of the elastic wave equation is given by the following set of relations (Virieux, 1986):

$$\rho(\vec{x})\frac{\partial v_x(\vec{x},t)}{\partial t} - \frac{\partial \sigma_{xx}(\vec{x},t)}{\partial x} - \frac{\partial \sigma_{xz}(\vec{x},t)}{\partial z}, = S_{vx}(\vec{x},t) \tag{1}$$

$$\rho(\vec{x})\frac{\partial v_z(\vec{x},t)}{\partial t} - \frac{\partial \sigma_{xz}(\vec{x},t)}{\partial x} - \frac{\partial \sigma_{zz}(\vec{x},t)}{\partial z} = S_{vz}(\vec{x},t), \tag{2}$$

$$\frac{\partial \sigma_{xx}(\vec{x},t)}{\partial t} - [\lambda+2\mu](\vec{x})\frac{\partial v_x(\vec{x},t)}{\partial x} - \lambda(\vec{x})\frac{\partial v_z(\vec{x},t)}{\partial z} = S_{xx}(\vec{x},t), \tag{3}$$

$$\frac{\partial \sigma_{zz}(\vec{x},t)}{\partial t} - \lambda(\vec{x})\frac{\partial v_x(\vec{x},t)}{\partial x} - [\lambda+2\mu](\vec{x})\frac{\partial v_z(\vec{x},t)}{\partial z} = S_{zz}(\vec{x},t), \tag{4}$$

$$\frac{\partial \sigma_{xz}(\vec{x},t)}{\partial t} - \mu(\vec{x})\left[\frac{\partial v_x(\vec{x},t)}{\partial z} - \frac{\partial v_x(\vec{x},t)}{\partial z}\right] = S_{xz}(\vec{x},t), \tag{5}$$

where λ, and μ are the Lamé parameters, ρ is density, v_x, and v_z are the particle velocities, and σ_{xx}, σ_{zz}, and σ_{xz} are the propagated stresses. The variables on the right-hand side of these equations represent the forcing terms. As shown by Alves and Biondi (2016), we can rewrite these equations as a non-linear operator:

$$\vec{d} = f(\vec{m}), \tag{6}$$

where $\vec{d} = [v_x\ v_z\ \sigma_{xx}\ \sigma_{zz}\ \sigma_{xz}]^T$, and $\vec{m} = [\lambda\ \mu\ \rho]^T$ define our data and model vectors, respectively. In real seismic acquisition only the hydrostatic pressure and particle velocities are recorded. Therefore, we apply a linear transformation to the data vector to simulate a real experiment:

$$\vec{d}_{obs} = \mathbf{R}\vec{d}, \tag{7}$$

where the matrix $\mathbf{R}$ is defined as follows:

$$\begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}, \tag{8}$$

and whose effect is to average the normal stresses and extract the particle velocities from the data vector. The choice of model parametrization influences the inversion results because of the presence of parameter crosstalk (Operto et al., 2013). We choose to parametrize our model space with the vector $\vec{m}' = [V_p \; V_s \; \rho]^T$ that contains the wave propagation velocities as opposed to elastic parameters. This change of variables introduces the following non-linear transformation:

$$\vec{m} = g(\vec{m}') = \begin{bmatrix} (V_p^2 - 2V_s^2)\rho \\ V_s^2 \rho \\ \rho \end{bmatrix}, \tag{9}$$

where V_p and V_s are the compressional- and shear-wave propagation velocities, respectively. Given the previous equations we define our FWI objective function as:

$$\phi(\vec{m}') = \frac{1}{2} \left\| \mathbf{R} f\left(g(\vec{m}')\right) - \vec{d}_{obs} \right\|_2^2 = \frac{1}{2} \left\| \vec{r} \right\|_2^2, \tag{10}$$

where $\vec{d}_{obs}$ is the observed multi-component data, $\vec{r}$ is the residual vector, and we parametrized the modeling operator in terms of the $\vec{m}'$ vector. To minimize this function we compute the gradient of equation 10 that can be written as follows:

$$\nabla \phi = \mathbf{G}^T \mathbf{F}^T \mathbf{R}^T \vec{r}, \tag{11}$$

where $\mathbf{G}^T$ and $\mathbf{F}^T$ are the linearized adjoint operators of equations 9 and 6, respectively. The linearized adjoint operator $\mathbf{F}^T$ can be found using the adjoint state method (Fichtner, 2010), and the Jacobian matrix of the non-linear transformation of equation 9 is given by:

$$\mathbf{G} = \begin{bmatrix} 2V_p\rho & -4V_s\rho & V_p^2 - 2V_s^2 \\ 0 & 2V_s\rho & V_s^2 \\ 0 & 0 & 1 \end{bmatrix}. \tag{12}$$

Using equation 11 we apply any gradient-based optimization algorithm, such as non-linear conjugate gradient (CG) (Fletcher and Reeves, 1964). To improve the convergence rate of any optimization method we can apply an approximated inverse Hessian matrix to precondition the gradient. In fact, this operation would approximate a Newton's optimization step. The sparse structure of the Hessian enables us to estimate its elements by applying this matrix to impulses in the model space and interpolating for the unknown values (Tang and Lee, 2015). In this study we make use of the Gauss-Newton Hessian. For the objective function in equation 10 this matrix $\mathbf{H}_{GN}$ takes the following form:

$$\mathbf{H}_{GN} = \mathbf{G}^T \mathbf{F}^T \mathbf{R}^T \mathbf{R} \mathbf{F} \mathbf{G}, \tag{13}$$

where $\mathbf{F}$ is the Born operator that maps perturbations in the model space into data perturbations.

RESULTS AND DISCUSSION

We generate a 2D complex subsurface model using the model builder software proposed by Clapp (2014). Figure 1 shows this model in terms of compressional-wave velocity, shear-wave velocity, and density. In such model, we generate multi-component data by placing 50 explosive sources at the surface spaced by 100 m. As source signature we employ a Ricker wavelet with dominant frequency of 20 Hz. The hydrophones and geophones are also positioned at the surface but with an interval of 10 m. Figure 2 displays a representative shot multi-component record.

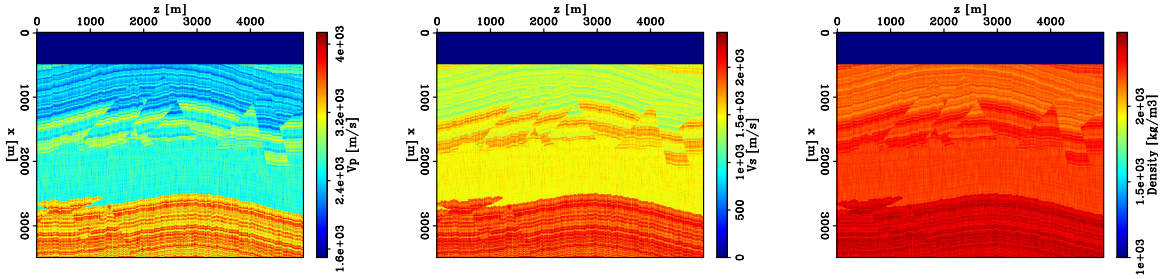

Figure 1: True subsurface model. Left panel: Compressional-wave velocity. Central panel: Shear-wave velocity. Right panel: Subsurface density. [**ER**]

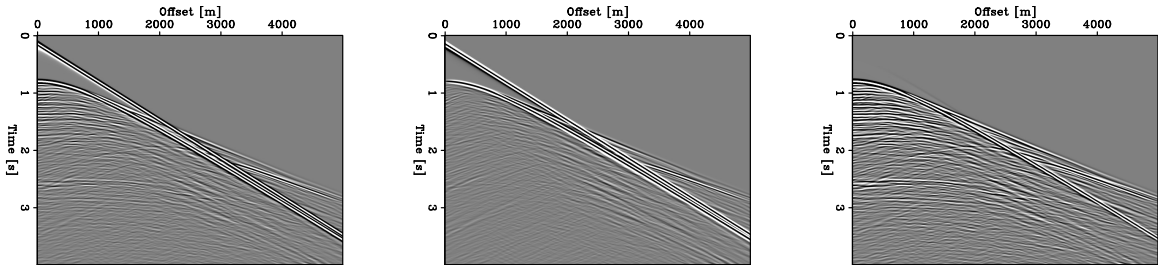

Figure 2: Multi-component data recorded for an explosive source placed on the surface at x = 0 m. Left panel: pressure component. Central panel: horizontal particle velocity component. Right panel: particle velocity component. [**CR**]

The initial model parameters used to start the FWI problem is constructed by smoothing the ones displayed in Figure 1 (Figure 3). To avoid local minima and increase the attraction basin of the global minimum, we follow a multi-scale approach (Bunks et al., 1995). We start the inversion with a bandwidth of maximum frequency of 5 Hz, and we increase the bandwidth by intervals of 5 Hz up to 20 Hz. The last inverted model from one band is going to be our initial subsurface for the next one. For each frequency band we run 40 iterations of non-linear CG algorithm.

We compare an optimization result obtained without employing any preconditioning, and a different one where an approximated Gauss-Newton Hessian inverse is estimated for each frequency band and is used to construct a preconditioner. To estimate this matrix, we apply the matrix of equation 13 to twelve impulses placed in the subsurface and spaced by 500 m and 1000 m along the z-axis and x-axis, respectively (Figure 4). From these applications we extract the main diagonal elements and

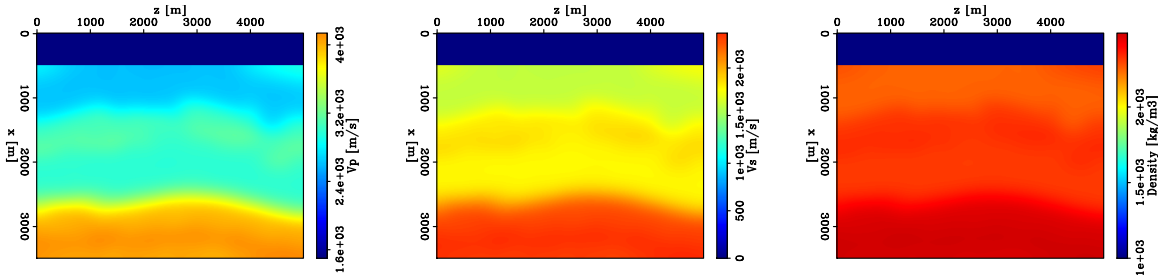

Figure 3: Starting FWI model created by smoothing the true model shown in Figure 1. Left panel: Compressional-wave velocity. Central panel: Shear-wave velocity. Right panel: Subsurface density. [**ER**]

linearly interpolate the unknown Hessian values. We then create an approximated Gauss-Newton Hessian inverse that is used to precondition any computed FWI gradient. This preconditioner is estimated only once for each frequency band before starting the inversion. This approach assumes that this weighting is not changing as the FWI is modifying the model. This assumption is reasonable since the model perturbations introduced by the FWI scheme will not drastically change the Gauss-Newton Hessian matrix.

Figure 5 shows the elastic FWI results. For all parameter classes, the inverted models are similar in the shallow portion of the subsurface, meaning both FWI optimizations have likely converged to the same minimum. However, in the deeper layers, they differ in terms of parameter resolution. In fact, when the optimization is preconditioned the inverted parameters present more structural features. In addition, a low-density layer on top of the interface approximately at 2700 m does not affect the inverted density model. In terms of objective function values, both inversion results have similar behavior. From Figure 6 we observe a modest improvement in convergence rate when the problem is preconditioned with a Gauss-Newton main-diagonal approximation, especially as we increase the frequency content in the inversion. In this test, the effect the preconditioner is to mostly decrease the model crosstalk by properly scaling the gradients of the simultaneously inverted parameter classes.

CONCLUSIONS

We discuss how to derive the Gauss-Newton Hessian approximation of elastic multicomponent FWI as a series of linear operators when wave velocities and density parameterize the inverse problem. We describe how to estimate these matrix elements from applications to model vectors containing sparse impulses. Using linear interpolation we compute the missing matrix elements.

On a complex 2D synthetic model we show the use of this estimated matrix to precondition the FWI problem when multiple parameters are inverted simultaneously. We demonstrate that a simple main-diagonal approximation already provides a mod-

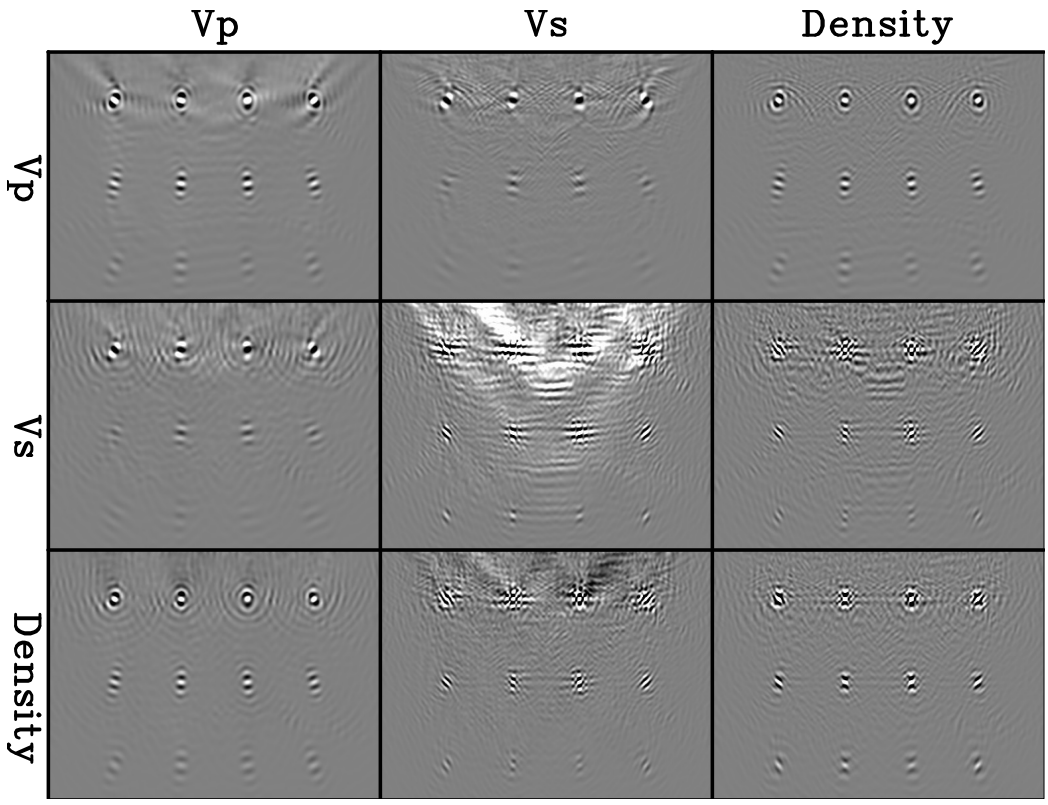

Figure 4: Application of the Gauss-Newton Hessian for the 15 Hz frequency band to spikes positioned in the subsurface. The top labels indicate the parameter image obtained by applying the Hessian matrix to impulses in the parameter class indicated by the left label. The diagonal panels enable us to estimate the main diagonal of the Gauss-Newton Hessian matrix. In these panels the water layer has been removed. [**CR**]

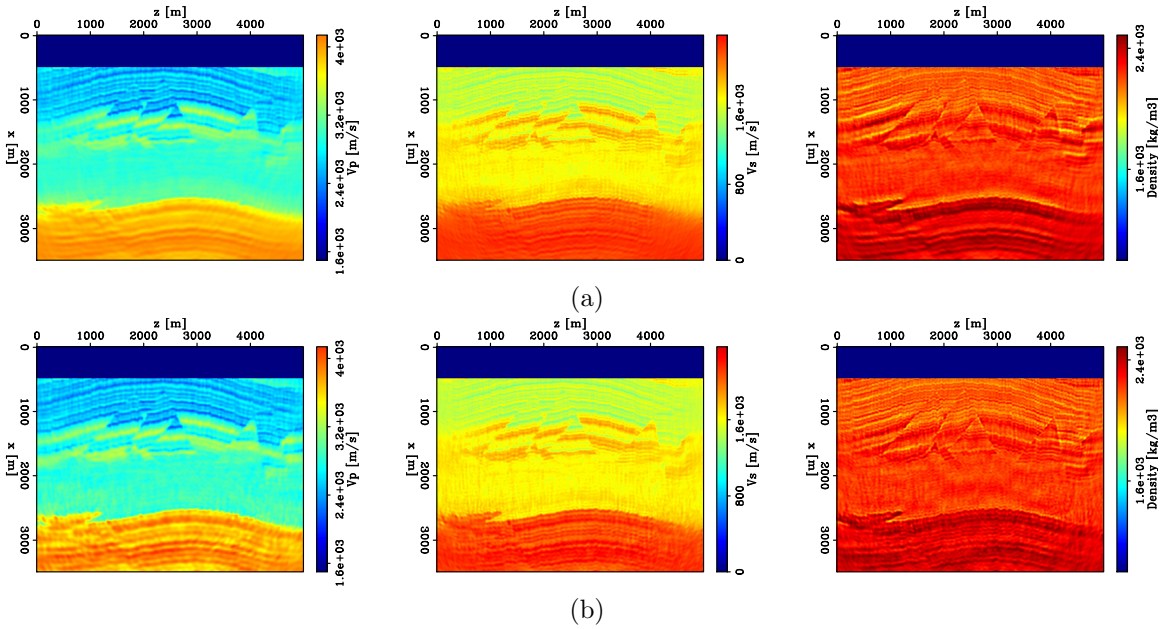

Figure 5: Inverted FWI model comparison. (a) Inverted model without preconditioning. (b) Inverted model using approximated Gauss-Newton inverse. [**CR**]

erate convergence improvement and properly scales the FWI gradients. In fact, this scaling removes inversion artifacts in the final inverted model. Future works will explore the effect of introducing more elements when approximating the Hessian inverse as preconditioning matrix.

REFERENCES

Albertin, U., P. Shen, A. Sekar, T. Johnsen, C. Wu, K. Nihei, and K. Bube, 2016, 3D orthorhombic elastic full-waveform inversion in the reflection domain from hydrophone data: SEG Technical Program Expanded Abstracts 2016, 1094–1098, Society of Exploration Geophysicists.

Alkhalifah, T., 2015a, Conditioning the full-waveform inversion gradient to welcome anisotropy: Geophysics, **80**, R111–R122.

———, 2015b, Scattering-angle based filtering of the waveform inversion gradients: Geophysical Journal International, **200**, 363–373.

Alves, G. and B. Biondi, 2016, Imaging condition for elastic reverse time migration: SEG Technical Program Expanded Abstracts 2016, 4173–4177, Society of Exploration Geophysicists.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Clapp, R., 2014, Synthetic model building using a simplified basin modeling approach: SEP-Report, **155**, 143–150.

Deuzeman, A. and R.-E. Plessix, 2015, Block-diagonal approximation of the Hessian

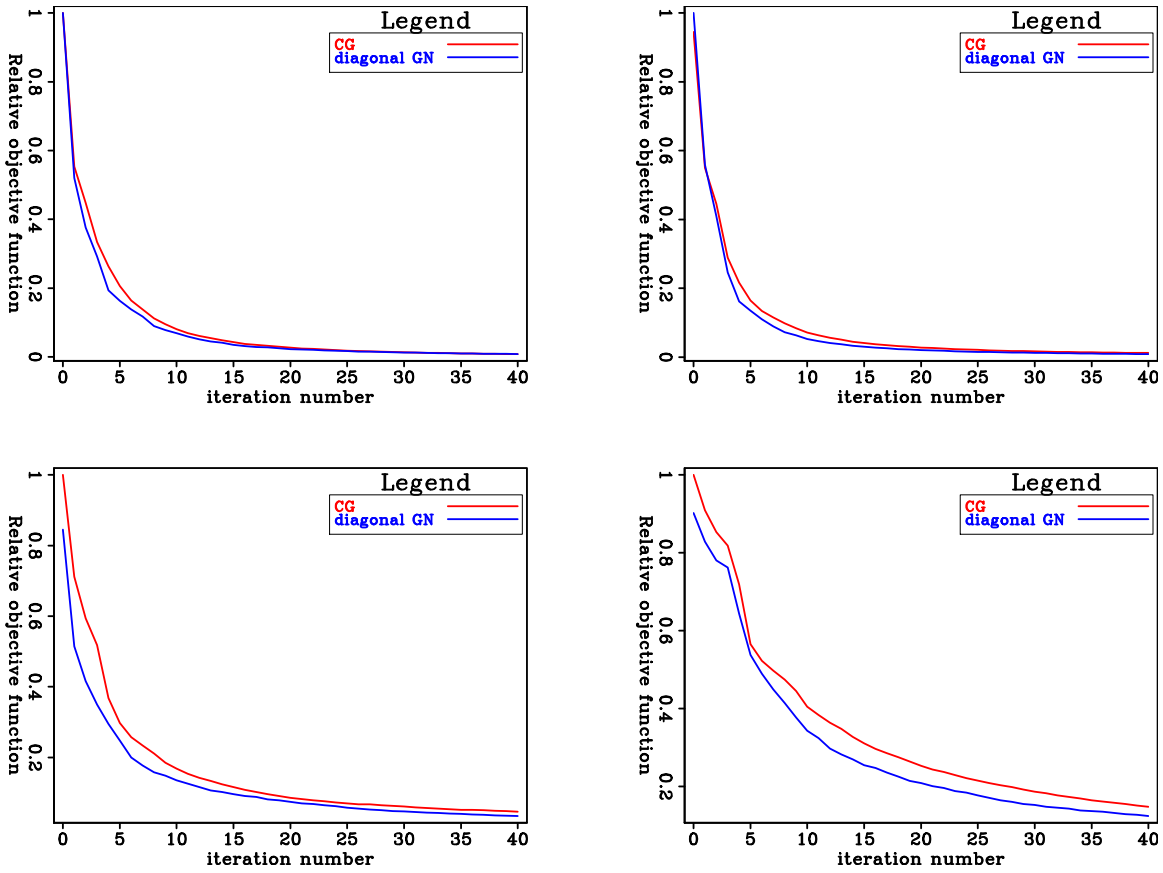

Figure 6: Relative objective function comparison between CG (red curve) and preconditioned CG (blue curve) for maximum frequency content of 5 (top-left), 10 (top-right), 15 (bottom-left), and 20 Hz (bottom-right) used during the inversion. [**CR**]

for multi-parameter FWI: 12th SIAM Conference on Mathematical and Computational Issues in Geosciences, Expanded Abstracts, MS2, Society of Industrial and Applied Mathematics.

Fichtner, A., 2010, Full seismic waveform modelling and inversion: Springer Science & Business Media.

Fletcher, R. and C. M. Reeves, 1964, Function minimization by conjugate gradients: The computer journal, **7**, 149–154.

Forgues, E. and G. Lambaré, 1997, Parameterization study for acoustic and elastic ray plus born inversion: Journal of Seismic Exploration, **6**, 253–277.

Huang, Y., R. Nammour, and W. Symes, 2016, Flexibly preconditioned extended least-squares migration in shot-record domain: Geophysics, **81**, S299–S315.

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

Korta, N., A. Fichtner, and V. Sallarčs, 2013, Block-diagonal approximate hessian for preconditioning in full waveform inversion: Presented at the 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013.

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

Tang, Y., 2008, Wave-equation hessian by phase encoding: SEG Technical Program Expanded Abstracts 2008, 2201–2205, Society of Exploration Geophysicists.

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

ten Kroode, F., 2012, A wave-equation-based Kirchhoff operator: Inverse Problems, **28**, 115013.

Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, 889–901.