Dissection of the full-waveform inversion Hessian

Biondo Biondi, Ettore Biondi, Musa Maharramov, and Yinbin Ma

ABSTRACT

We analyze the FWI Hessian and determine that is the combination of two well-understood operators in seismic imaging: 1) the Gauss-Newton Hessian and 2) a differential WEMVA operator. We illustrate this insight with several numerical examples of applying and inverting the different components of the Hessian operator to images originating from the waveform inversion of a simple synthetic dataset. We then discuss possible applications of the theory to making FWI more robust with respect to incorrect starting models, and to speeding-up the application of target-oriented fully non-linear seismic inversion.

INTRODUCTION

Improving the robustness and the convergence rate of full waveform inversion (FWI) is an important goal for any FWI algorithm. It becomes particularly important when we go beyond simple constant-density acoustic inversion and we aim at estimating several images representing different parameters, such as elastic properties, anisotropic parameters and attenuation parameters (Alves, 2015; Biondi and O'Reilly, 2015; Shen, 2015).

Newton-like methods that exploit the information contained in the Hessian matrix of the objective function (Pratt et al., 1998) are attractive and have been recently the subject of much research (Tang and Lee, 2010; Fichtner, 2011; Métivier et al., 2012; Baumstein, 2014; Deuzeman and Plessix, 2015). Pratt et al. (1998) presented an efficient frequency-domain method for evaluating the application of the Hessian to a model-perturbations vector. Epanomeritakis et al. (2008) presented an equivalent method for time-domain propagation. In this paper we start from the description of the time-domain method for applying the Hessian to a model-perturbations vector presented by Fichtner and Trampert (2011) and Fichtner (2011).

In the following section we analyze the full FWI Hessian and show that it is the composition of well-understood operators in seismic imaging; that is, the Gauss-Newton Hessian and the WEMVA operator. This new insight has the potential of leading to more robust and computationally efficient waveform-inversion algorithms. We discuss these potential applications in the discussion section of the paper, after we illustrate the properties of each component of the Hessian operator by showing its application to a simple FWI problem.

DISSECTION OF THE FULL FWI HESSIAN OPERATOR

In conventional amplitude-based full waveform inversion (FWI), we minimize the following objective function:

$$\phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{d}_m(\mathbf{m}) - \mathbf{d}_o\|_2^2, \tag{1}$$

where $\mathbf{m}$ is the vector of model parameters, $\mathbf{d}_o$ is the vector of recorded data, and $\mathbf{d}_m(\mathbf{m})$ is the vector of modeled data. The modeled data are given by:

$$\mathbf{d}_m(\mathbf{m}) = \mathbf{R}\mathbf{u}(\mathbf{m}), \tag{2}$$

in which $\mathbf{R}$ is a linear operator that extracts the vector of modeled data, $\mathbf{d}_m$, from the vector of propagated wavefield, $\mathbf{u}$. In the following we assume that this wavefield is the solution of the following scalar wave equation:

$$\mathbf{L}\left(\mathbf{m},\mathbf{u}\right) = \left[m\frac{\partial^2}{\partial t^2} - \nabla^2\right] u(\mathbf{x},t,m) = s(\mathbf{x},t), \tag{3}$$

where m represents the slowness squared, s is the source term, and t and $\mathbf{x}$ are the temporal and spatial coordinates, respectively. However, the insights on the full Hessian presented in this paper are valid and applicable also when the data are modeled using more complex wave equations than the one presented in equation 3.

The objective function 1 aims to minimize in a least-squares sense the data residuals, defined as:

$$\mathbf{\Delta d} = \mathbf{d}_m(\mathbf{m}) - \mathbf{d}_o. \tag{4}$$

The Hessian operator has an important role when solving the optimization problem of minimizing the objective function 1 by a Newton algorithm. If $\nabla\phi$ is the gradient vector, an improved search direction $\mathbf{\Delta m}$ can be obtained by solving the Newton system

$$\mathbf{H}\mathbf{\Delta m} = -\nabla\phi, \tag{5}$$

where $\mathbf{H}$ is the Hessian matrix. Because of the large scale of the FWI problem we seldom build the Hessian matrix explicitly, and never solve the Newton system directly. On the contrary, we invert the Hessian iteratively using an iterative method like conjugate-gradient. Furthermore, we do not iterate the inversion of $\mathbf{H}$ to full-convergence, but we stop after a small number of iterations. This approximate Newton algorithm is often referred as *truncated Newton method*.

To invert $\mathbf{H}$ iteratively we need only to be able to compute its application to an arbitrary model perturbations $\delta\mathbf{m}$. To perform this task we can precompute the Hessian matrix and perform several matrix-vector operations, or devise an algorithm that evaluates the application of $\mathbf{H}$ to a vector by performing wave propagations and correlations. The latter way is conceptually similar to the one employed to compute the

FWI gradient using adjoint-state methods, and is the one that we will discuss in this section. In practice, hybrid schemes (i.e. methods that rely on precomputing some Hessian components and applying the adjoint-state methods for other components) can be actually the most computationally efficient solution.

If we start the iterative inversion of the Hessian matrix from a zero model-perturbations vector, the first search direction, $\Delta\mathbf{m}_1$, of a truncated Newton algorithm is equal to,

$$\Delta\mathbf{m}_1 = -\mathbf{H}^{'}\nabla\phi = -\mathbf{H}\nabla\phi, \tag{6}$$

where in equation 6 we use the Hermitian properties of the Hessian matrix.

In the numerical examples section we will show both the results of approximately solving the Newton system (equation 5), as well as computing the search direction $\Delta\mathbf{m}_1$ using equation 6.

The Hessian matrix is the second-order derivative of the objective function with respect to each model parameter; it can be expressed as follows:

$$\nabla_m\nabla_m\phi(\mathbf{m}) = (\nabla_u\nabla_u\phi\nabla_m\mathbf{u})\cdot\nabla_m\mathbf{u} + (\nabla_u\phi)\cdot\nabla_m\nabla_m\mathbf{u}, \tag{7}$$

where the first term on the right-hand side of the equation represents the Gauss-Newton Hessian.

As shown in Fichtner (2011), the application of the full FWI Hessian to a model perturbation, $\delta\mathbf{m}$, can be written as sum of three kernels applied to $\delta\mathbf{m}$ as:

$$\mathbf{H}(\mathbf{m}_0)\delta\mathbf{m} = [\mathbf{K}^a(\mathbf{m}_0) + \mathbf{K}^b(\mathbf{m}_0) + \mathbf{K}^c(\mathbf{m}_0)]\delta\mathbf{m}, \tag{8}$$

where $\mathbf{m}_0$ is the model around which we are linearizing the nonlinear objective function and the three kernels are:

$$\mathbf{K}^a(\mathbf{m}_0)\delta\mathbf{m} = \left[\int_t \mathbf{u}_2^{'}\cdot\nabla_m\mathbf{L}\left(\mathbf{m}_0,\mathbf{u}_0\right)\ dt\right]\delta\mathbf{m}, \tag{9}$$

$$\mathbf{K}^b(\mathbf{m}_0)\delta\mathbf{m} = \left[\int_t \mathbf{u}_1^{'}\cdot\nabla_m\nabla_u\mathbf{L}\left(\mathbf{m}_0,\mathbf{u}_0\right)\nabla_m\mathbf{u}\ dt\right]\delta\mathbf{m}, \tag{10}$$

$$\mathbf{K}^c(\mathbf{m}_0)\delta\mathbf{m} = \left[\int_t \mathbf{u}_1^{'}\cdot\nabla_m\nabla_m\mathbf{L}\left(\mathbf{m}_0,\mathbf{u}_0\right)\ dt\right]\delta\mathbf{m}, \tag{11}$$

where $\int_t$ denotes time integration, $\mathbf{u}_0$ is the background wavefield that is the solution of:

$$\mathbf{L}(\mathbf{m}_0,\mathbf{u}_0) = s(\mathbf{x},t), \tag{12}$$

and $\mathbf{u}_1^{'}$ and $\mathbf{u}_2^{'}$ are the *primary adjoint* wavefield and *secondary adjoint* wavefield, respectively.

In equation 3 we parametrized the wave equation in terms of slowness squared; therefore, we can drop the third kernel of the full Hessian, $\mathbf{K}^c$, because $\nabla_m\nabla_m\mathbf{L}$ is zero. The kernel $\mathbf{K}^c$ would not be zero if more complex wave-equation were employed

to model the data and the medium parameters were present non-linearly in the equation (e.g. Vp and Vs in a fully elastic wave equation). However, its evaluation would follow a flow similar the one used to evaluate $\mathbf{K}^b$.

The first step to interpret the two remaining kernels, $\mathbf{K}^a$ and $\mathbf{K}^b$, is the analysis of the two adjoint wavefields, $\mathbf{u}_1^{'}$ and $\mathbf{u}_2^{'}$. The primary adjoint wavefield, $\mathbf{u}_1^{'}$, is the solution of:

$$\nabla_u^{'} \mathbf{L}^{'} \mathbf{u}_1^{'} = -\nabla_u \phi. \tag{13}$$

Because $\mathbf{L}$ is linear with respect to the wavefield, and $\nabla_u \phi = \mathbf{\Delta d}$, equation 13 simplifies into

$$\mathbf{L}^{'}(\mathbf{m}_0, \mathbf{u}_1^{'}) = -\mathbf{\Delta d}. \tag{14}$$

The wavefield $\mathbf{u}_1^{'}$ is obtained by backpropagating the opposite of the data residual with the background model $\mathbf{m}_0$. It is the same wavefield that is computed to evaluate the gradient of objective function 1, and thus does not need to be recomputed when applying the Hessian operator.

The secondary adjoint wavefield, $\mathbf{u}_2^{'}$, is the solution of the following wave propagation problem:

$$\mathbf{L}^{'}(\mathbf{m}_0, \mathbf{u}_2^{'}) = -\left[\nabla_u \nabla_u \phi \nabla_m \mathbf{u} + \nabla_u \nabla_m \mathbf{L}^{'} \mathbf{u}_1\right]. \tag{15}$$

Since there are two source terms on the right hand side of equation 15, the secondary adjoint wavefield can be split into two components, $\mathbf{u}_{21}^{'}$ and $\mathbf{u}_{22}^{'}$, each solution of a backpropagation problem; that is:

$$\nabla_u \mathbf{L}^{'} \mathbf{u}_{21}^{'} = -\nabla_u \nabla_u \phi \nabla_m \mathbf{u}, \tag{16}$$

$$\nabla_u \mathbf{L}^{'} \mathbf{u}_{22}^{'} = -\nabla_u \nabla_m \mathbf{L}^{'} \mathbf{u}_1. \tag{17}$$

Consequently, the kernel $\mathbf{K}^a$ can be split in the sum of two kernels:

$$\mathbf{K}^{a1}(\mathbf{m}_0)\delta\mathbf{m} = \left[\int_t \mathbf{u}_{21}^{'} \cdot \nabla_m \mathbf{L}\left(\mathbf{m}_0, \mathbf{u}_0\right)\ dt\right] \delta\mathbf{m}, \tag{18}$$

$$\mathbf{K}^{a2}(\mathbf{m}_0)\delta\mathbf{m} = \left[\int_t \mathbf{u}_{22}^{'} \cdot \nabla_m \mathbf{L}\left(\mathbf{m}_0, \mathbf{u}_0\right)\ dt\right] \delta\mathbf{m}. \tag{19}$$

Next we interpret the first one, $\mathbf{K}^{a1}$. In equation 16, the derivative of the wavefield with respect of the model parameters can be written as the solution of the forward first-order Born scattering problem (Fichtner, 2011); that is:

$$\mathbf{L}(\mathbf{m}_0, \delta\mathbf{u}) = -\nabla_m \mathbf{L}(\mathbf{m}_0, \mathbf{u}_0). \tag{20}$$

Furthermore, in equation 16, $\nabla_u \nabla_u \phi$ is simply $\mathbf{R}^{'}\mathbf{R}$, which means extracting and reinjecting $\delta\mathbf{u}$ at the receiver positions.

Equation 16 can thus be rewritten as

$$\mathbf{L}^{'}(\mathbf{m}_0, \mathbf{u}_{21}^{'}) = \mathbf{R}^{'}\mathbf{R}\delta\mathbf{u}, \tag{21}$$

where $\delta\mathbf{u}$ is given by equation 20. After cross-correlation with the second time derivative of the forward propagated background wavefield ($\nabla_m\mathbf{L}(\mathbf{m}_0, \mathbf{u}_0)$), as indicated by equation 18, the application of $\mathbf{K}^{a1}$ to a model perturbation $\delta\mathbf{m}$ is equivalent to the application of the forward first-order Born modeling operator followed by the application of its adjoint. In other words, it corresponds to Born modeling followed by migration. This kernel corresponds to the Gauss-Newton component of the full Hessian. From now on, we will denote it as $\mathbf{H}_{\text{GN}}$.

Next we look at the interpretation of $\mathbf{K}^{a2}$ and we analyze the source term in equation 17. This term corresponds to the first-order Born scattering of $\mathbf{u}_1^{'}$, which is the the backpropagated wavefield generated by injecting the data residuals at the receivers (equation 14). The wavefield $\mathbf{u}_{22}^{'}$ is thus obtained by: 1) computing the primary adjoint wavefield, $\mathbf{u}_1^{'}$, 2) scattering $\mathbf{u}_1^{'}$ at $\delta\mathbf{m}$, and then, 3) backpropagating the scattered wavefield backward in time. The kernel $\mathbf{K}^{a2}$ is then evaluated by cross-correlating $\mathbf{u}_{22}^{'}$ with the the second time derivative of the forward propagated background wavefield ($\nabla_m\mathbf{L}(\mathbf{m}_0, \mathbf{u}_0)$) as indicated by equation 19.

The kernel $\mathbf{K}^{a2}$ is similar to the *source* side of the adjoint of the conventional WEMVA operator, with an important difference (Sava and Vlad, 2008). The wavefield $\mathbf{u}_{22}^{'}$ is generated by injecting the data residuals, not the recorded data, as is done for the conventional WEMVA operator. However, because the operator is linear with respect to the injected wavefields, it can also be interpreted as the difference of two WEMVA operators: the first, $\mathbf{W}_s^m$, is obtained by injecting the modeled data $\mathbf{d}_m$ at the receivers, whereas the second, $\mathbf{W}_s^o$, is obtained by injecting the observed data $\mathbf{d}_o$. This second operator is exactly equivalent to the adjoint of the conventional WEMVA operator.

The next, and final, step of our analysis is to interpret the kernel $\mathbf{K}^b$, as defined by equation 10. We first notice that because of the linearity of $\mathbf{L}$ with respect to the wavefield we can write the following identity,

$$\nabla_m\nabla_u\mathbf{L}\left(\mathbf{m}_0, \mathbf{u}_0\right) = \nabla_m\mathbf{L}\left(\mathbf{m}_0, \mathbf{u}_0\right). \tag{22}$$

Then we notice that, as for equation 16, the derivative of the wavefield with respect to the model parameters is the solution of the forward first-order Born scattering problem expressed in equation 20. Therefore, the applications of kernel $\mathbf{K}^b$ to the model perturbation $\delta\mathbf{m}$ is evaluated by cross-correlating the forward scattered wavefield $\delta\mathbf{u}$ with the second time derivative of the backward propagated primary adjoint wavefield $\mathbf{u}_1^{'}$. This kernel $\mathbf{K}^b$ is similar to the *receiver* side of the adjoint of the conventional WEMVA operator. As discussed above for $\mathbf{K}^{a2}$, the operator $\mathbf{K}^b$ can also be interpreted as the difference of two WEMVA operators: the first, $\mathbf{W}_r^m$, is obtained by injecting the modeled data $\mathbf{d}_m$ at the receivers, whereas the second, $\mathbf{W}_r^o$, is obtained by injecting the observed data $\mathbf{d}_o$.

In conclusions the full Hessian operator $\mathbf{H}$ is the composition of familiar operators, as follows

$$\mathbf{H} = \mathbf{H}_{\rm GN} + (\mathbf{W}_s^m - \mathbf{W}_s^o) + (\mathbf{W}_r^m - \mathbf{W}_r^o) = \mathbf{H}_{\rm GN} + (\mathbf{W}^m - \mathbf{W}^o) = \mathbf{H}_{\rm GN} + \mathbf{H}_{\rm W}, \quad (23)$$

where $\mathbf{W}^m$ and $\mathbf{W}^o$ are the WEMVA operators obtained by injecting the modeled and observed data, respectively; $\mathbf{H}_{\rm W}$ is the difference between $\mathbf{W}^m$ and $\mathbf{W}^o$. We will refer to this difference as the *differential WEMVA* component of the full Hessian.

If the wave-equation modeling operator ($\mathbf{L}$) is sufficiently accurate to model all the important features present in the recorded data, as the inversion process converges to a model closer to the true model, the modeled data becomes closer to the observed data. In this situation, the corresponding WEMVA operators $\mathbf{W}^m$ and $\mathbf{W}^o$ become closer to each other; their difference becomes negligible and the full Hessian is equivalent to the simple Gauss-Newton Hessian. However, when the wave-equation modeling operator ($\mathbf{L}$) is too simplistic to model correctly important wave phenomena (e.g. elastic, multiples, attenuation) the modeled data does not converge to the recorded data even close to the global minimum of objective function 1, and consequently the full Hessian may never converges towards the Gauss-Newton Hessian. The implications of this lack of convergence of the full Hessian towards the Gauss-Newton Hessian for Newton methods are not obvious, and may deserve further studies.

Because the gradient is a linear function of the data injected at the surface, we can consider the gradient $\nabla\phi$ as the difference of two terms corresponding to the modeled data ($\nabla\phi^m$) and the observed data ($\nabla\phi^o$); that is,

$$\nabla\phi = (\nabla\phi^m - \nabla\phi^o)\,. \quad (24)$$

Then, following equation 6, the first search direction, $\Delta\mathbf{m}_1$, of a truncated Newton algorithm can be decomposed in 6 terms as follows:

$$\begin{aligned}
\Delta\mathbf{m}_1 &= -\mathbf{H}\nabla\phi = \\
&- \mathbf{H}_{\rm GN}\nabla\phi^m && (25)\\
&+ \mathbf{H}_{\rm GN}\nabla\phi^o && (26)\\
&- \mathbf{W}^m\nabla\phi^m && (27)\\
&+ \mathbf{W}^m\nabla\phi^o && (28)\\
&+ \mathbf{W}^o\nabla\phi^m && (29)\\
&- \mathbf{W}^o\nabla\phi^o. && (30)
\end{aligned}$$

In the numerical example section we will show examples each of these terms and discuss their roles.

Observations

Equation 23 shows the decomposition of the full FWI Hessian into two operators. We now discuss each term, and discuss its role in a truncated Newton FWI solution.

Gauss-Newton Hessian

The Gauss-Newton component of the Hessian ($\mathbf{H}_{\mathrm{GN}}$) describes the correlation between model parameters that are caused by the finite-resolution nature of the seismic imaging process (data-acquisition followed by migration). The resolution limits are determined by the finite frequency bandwidth and finite aperture of realistic seismic acquisitions, or in the case of multi-parameter inversion (anisotropic, elastic, visco-acoustic), the trade off between the parameters used to describe the unknown medium. Using an image-processing terminology, the application of $\mathbf{H}_{\mathrm{GN}}$ to a model perturbation is the convolution with non-stationary point-spread functions which takes into account the finite resolution of the seismic imaging system. Except in regions of the model that suffer from severe lack of illumination (i.e. shadow zones) the Gauss-Newton Hessian is diagonally dominant (or block diagonally dominant in case of multi-parameters inversion). However, $\mathbf{H}_{\mathrm{GN}}$ is also a singular matrix and thus its inversion must be regularized.

The application of the Gauss-Newton Hessian ($\mathbf{H}_{\mathrm{GN}}$) operator and its approximate inversion have been discussed in several SEP thesis (Valenciano, 2008; Tang, 2011; Ayeni, 2011) and in the wider literature. It is related to iterative linearized waveform inversion (Kuehl and Sacchi, 1999; Prucha et al., 2000) and least-squares migration (Jin et al., 1991; Cole and Karrenbach, 1992; Schuster, 1993; Nemeth et al., 1999).

In a single-parameter acoustic inversion the application of the approximate inverse of $\mathbf{H}_{\mathrm{GN}}$ to a migrated image improves resolution and the relative amplitudes among migrated reflectors. If appropriate regularization terms are added to the objective function in equation 1, inverting $\mathbf{H}_{\mathrm{GN}}$ can compensate for poor illumination caused by complex overburden (Clapp, 2005; Valenciano, 2008; Tang, 2011; Fletcher et al., 2012), or better equalize time-lapse images collected with different acquisition geometries (Ayeni, 2011).

When the subsurface is parametrized using more than one parameter, the inversion of $\mathbf{H}_{\mathrm{GN}}$ helps the resolution of the the trade off between model parameters and thus it can substantially improve convergence (Baumstein, 2014; Tang and Lee, 2015). For this application, $\mathbf{H}_{\mathrm{GN}}$ can be effectively approximated as a block-diagonal matrix since the modeling of the cross-talk between different parameters is its crucial contribution.

The application of $\mathbf{H}_{\mathrm{GN}}$ to a model-perturbations vector $\delta\mathbf{m}$ using the methodology introduced in this section requires two wavefield propagations: 1) forward propagation of the wavefield scattered by $\delta\mathbf{m}$ (equation 20), and 2) backward propagation of the scattered wavefield recorded and reinjected at the receivers (equation 21). Alternatively, because $\mathbf{H}_{\mathrm{GN}}$ is diagonally dominant, it can be approximated by a banded matrix and effectively precomputed and stored. The computation of the dominant elements of $\mathbf{H}_{\mathrm{GN}}$ can be approximated by phase-encoding methods (Tang, 2011; Deuzeman and Plessix, 2015) or by interpolation of the results of modeling and migrating isolated model perturbations (Fletcher et al., 2012; Tang and Lee, 2015).

Differential WEMVA component of the Hessian

The differential WEMVA component of the Hessian ($\mathbf{H}_{\mathrm{W}}$) describes the interaction between model parameters caused by multiple-scattering during wave propagation. In particular, the $\mathbf{H}_{\mathrm{W}}$ operator models the effects of second-order scattering. This second-order scattering can be both *short range* or *long range* (see Figure 4).

The long-range scattering is related to velocity errors that cause defocusing of the gradient. The application of ($\mathbf{H}_{\mathrm{W}}$) to the FWI gradient $\nabla\phi$ is equivalent to the gradient that we would obtain from a WEMVA method that aims to maximize the power of the stacked image, where the image is the FWI gradient itself, with a negative sign. The maximization of the image stack power is known to be prone to cycle-skipping problems when the background velocity is far away from the true velocity and the gradient is substantially misfocused. When the background model is far from the true one, the long-range component of the Hessian operator also cycle skips and points in the wrong direction (Figure 4(d)).

The short-range scattering of the WEMVA operator takes into account the interactions between nearby scatterers. It is related to amplitude and phase errors in the gradient caused by the first-order Born approximation. When the nearby scatterers are part of nearby reflectors, they may create internal multiples that are not modeled correctly by the first-order Born approximation. When these nearby scatterers are part of the same reflecting interface, neglecting their influence leads to errors in reflections amplitudes. For example, it is well-known that when we model reflection amplitudes from a planar interface using the first-order Born approximation we do not obtain the same amplitudes predicted by the Zoeppritz equations. Consequently, a linearized-Born inversion yields biased estimates of elastic properties at interfaces. To remove this bias, we must iterate the linearized-Born inversion in a fully non-linear algorithm.

The inclusion of the short-range scattering components of $\mathbf{H}_{\mathrm{W}}$ should improve elastic-properties estimates even without additional non linear iterations, and should speed up the convergence of a fully non-linear iterative inversion. Further analysis and tests will determine if the improved convergence justifies the additional computational cost.

SYNTHETIC EXAMPLE

To illustrate the effects of applying and inverting the Gauss-Newton and the full Hessian matrices we conducted several numerical tests based on a 2D synthetic dataset. To model the data we assumed a velocity model 3.5 km wide and 1 km deep. The background velocity was 1.5 km/s with a single reflecting layer at depth of 800 m (Figure 1). We modeled 88 regularly spaced sources and recorded the data with a fixed array of 351 receivers. The source interval was 40 m and the receiver interval was 10 m. To compute the Laplacian necessary for the wave propagation, we em-

ployed a 10th-order approximation; the minimum wavelength was 30 m and the grid spacing 5 m. As source signature we used a Ricker wavelet with central frequency of 10 Hz.

To invert the Hessian matrices we relied on the Intel MKL iterative sparse solvers (ISS) based on the reverse communication interface (RCI) (Dongarra et al., 1995). In our test we did not add a regularization term to the objective function. However, if necessary, we could regularize the inversion or apply the Leverberg-Marquardt technique to stabilize the inversion of the Hessian matrix (Pujol, 2007).

Figure 1 shows the velocity model used to model the data. Figure 2 shows one of the modeled shot gathers after muting the direct arrival. We run two iterations of FWI, with different choices of starting models and approximation of the Hessian matrix.

First-iteration results

Figure 3(a) shows the image (opposite of FWI gradient) obtained by assuming the correct background velocity, but no reflecting interface. The imaged reflector is well focused, though resolution suffers because of the effects of the finite-bandwidth wavelet. Figure 3(b) shows the image (opposite of FWI gradient) obtained by assuming a background velocity lower than the correct one by 2%. The image is not as well focused as in Figure 3(a) and the phase of the reflector is distorted. Finally, Figure 3(c) shows the image (opposite of FWI gradient) obtained by assuming a background velocity with a larger error (10%) than in the previous case; as expected, the phase of the reflector is more distorted than in the previous case.

Figure 4(a) is the search direction obtained by applying the full Hessian to the image obtained by assuming a background velocity too low by 2 % (image in Figure 3(b)). As explained in the previous section, this search direction ($\Delta\mathbf{m_1}$ in equation 6) is the first search direction of the iterative inversion of the Hessian matrix in a truncated Newton algorithm. Because we started from a constant model that did not produce any reflections, the modeled data after muting the first arrival is uniformly zero. Therefore, the application of the full Hessian to the image corresponds to only two terms out of the general 6 terms; that is, $\mathbf{H}_{\text{GN}}\nabla\phi^o$ (equation 26) and $-\mathbf{W}^o\nabla\phi^o$ (equation 30). Figure 4(b) is the search direction obtained by applying the Gauss-Newton Hessian (equation 26) to the same image as the results shown in Figure 4(a). In this case, the long-wavelength component is absent. We set the clip to be the same for displaying the two top panels in Figure 4; this clip was chosen to make visible the effects of applying to the gradient the WEMVA component of the Hessian easily visible.

Figure 4(c), shows the result of applying the WEMVA component of the Hessian operator that is, it is the difference between Figure 4(a) and Figure 4(b). The long wavelength of the model update points towards a negative slowness-squared update, and thus it points in the correct direction. In contrast, Figure 4(d) shows the result

of applying the WEMVA component of the Hessian operator to the image obtained by assuming a background velocity lower by 10% than the correct one (Figure 3(c)). The long wavelength of the model update points towards a positive slowness-squared update, and thus it points in the wrong direction.

Figure 5(a) shows the steepest-descent velocity update after a simple line search. Figure 5(b) shows the truncated Newton model update after iteratively inverting the Gauss-Newton Hessian and without line search (i.e. setting the step size to be equal to one). Figure 5(c) shows the truncated Newton model update after iteratively inverting the full Hessian and without line search. For both truncated Newton results the updated model shows a sharper image of the reflector than for the simple steepest descent result.

Figure 6(a) shows a vertical section of the three models shown in Figure 5. The updated model obtained by inverting the Gauss-Newton Hessian (green line) is similar to the model obtained by inverting the full Hessian (red line). There is only a small difference in the long-wavelengths component above the reflector, as shown in the zoomed-in plot displayed in Figure 6(b). The model updated by inverting the full Hessian has a small positive increase in the background velocity to correct for the initial error in velocity.

Figure 7(a) shows the normalized norm of the residuals (in model space) during the iterative inversion of the full Hessian. We did not regularize this inversion that started to diverge after 4 iterations. Figure 7(b) shows the normalized norm of the residuals during the iterative inversion of the Gauss-Newton Hessian. In this case the inversion converges even without regularization.

Figure 8 shows the same comparison shown in Figure 6(a), but corresponding to the case when the background velocity was correct. Now the inversion of the full Hessian is not trying to update the background model and the result is similar to the inversion of only the Gauss-Newton component of the Hessian. The applications of the inverse Hessians had the expected effect to reduce the side lobes in the image, and the updated models show a nicely focused reflector.

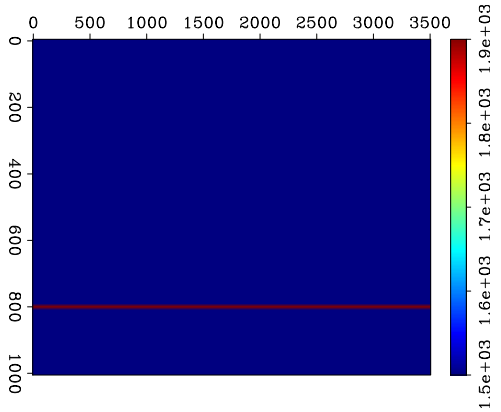


Figure 1: True velocity model with a single reflector (background 1.5 km/s perturbation .4 km/s) [**ER**]

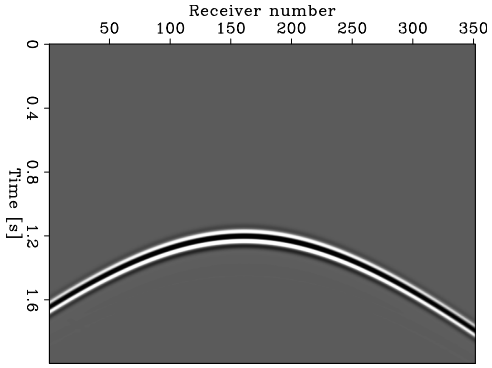


Figure 2: Common-shot gather of the observed data. Shot and receiver interval is 40 m and 10 m, respectively. [**ER**]

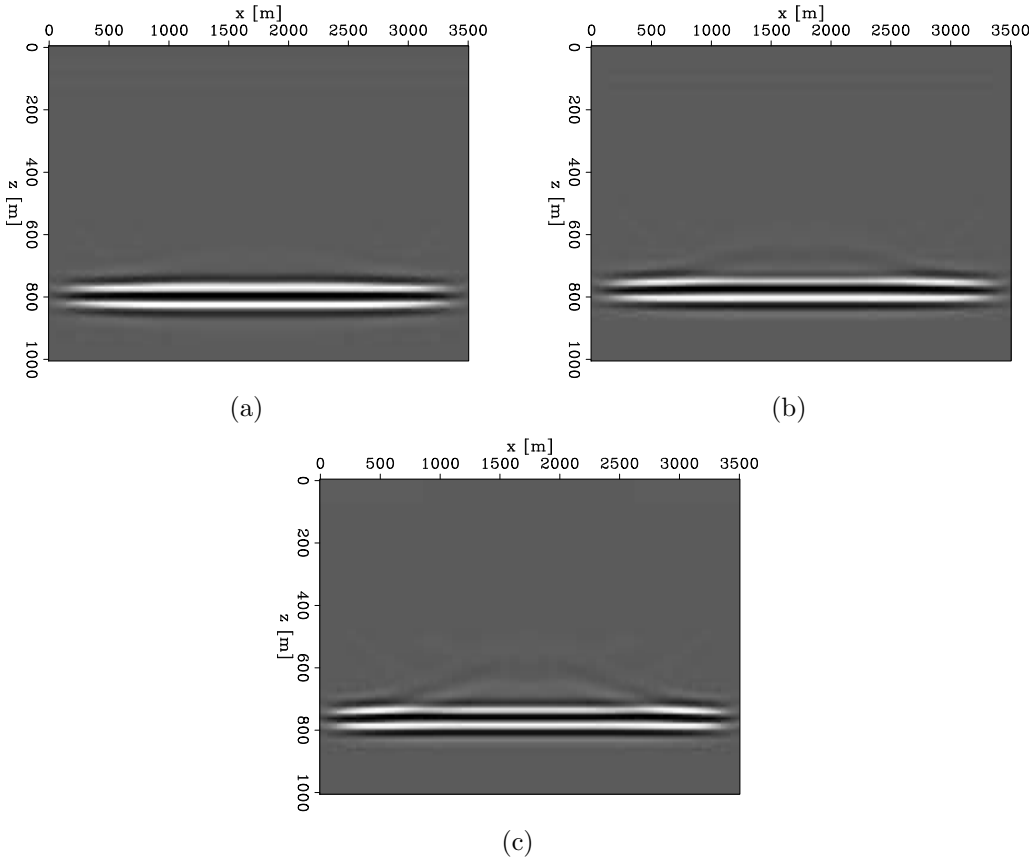


Figure 3: (a) Image computed with the correct background of 1.5 km/s. (b) Image obtained with a background with a negative velocity error of 2.0%. (c) Image obtained when the background error is -10%. [**ER**]

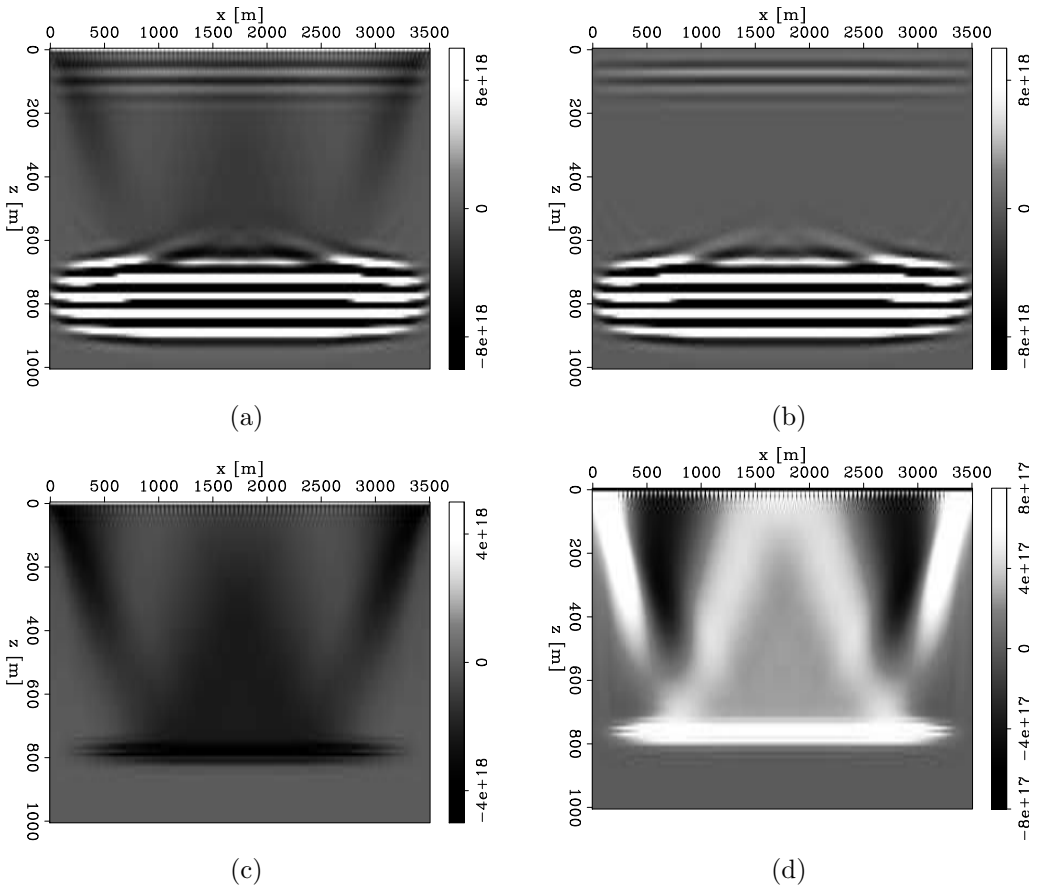
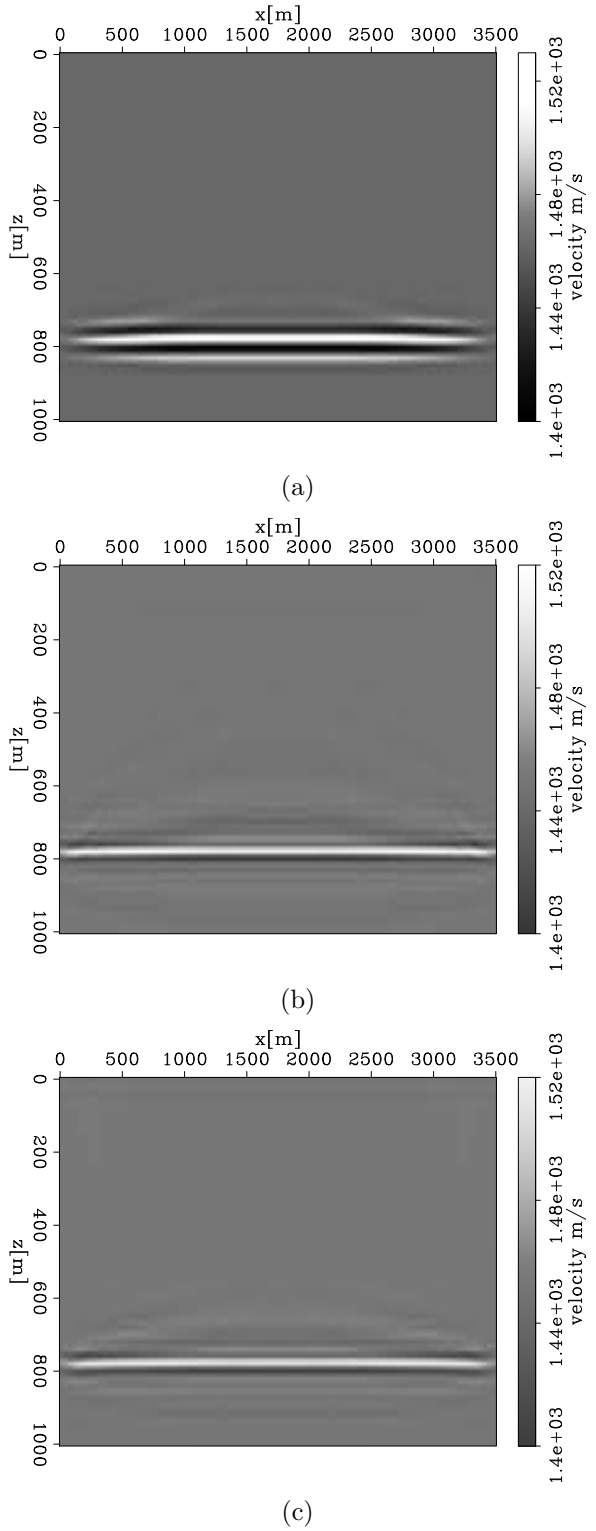


Figure 4: (a) Full and (b) Gauss-Newton Hessian applied to first search direction of Figure 3(b) (i.e., $-\mathbf{H}\nabla\phi$). (c) Difference between full and Gauss-Newton Hessians. Panel (a) and (b) have the same clipping value and they correspond to the first search direction during the Hessian inversion. (d) Hessian WEMVA component applied to the first search direction when the velocity error is -10% (Figure 3(c)). Because of the large velocity error in this case, the WEMVA search direction is pointing toward the incorrect velocity update. [**CR**]

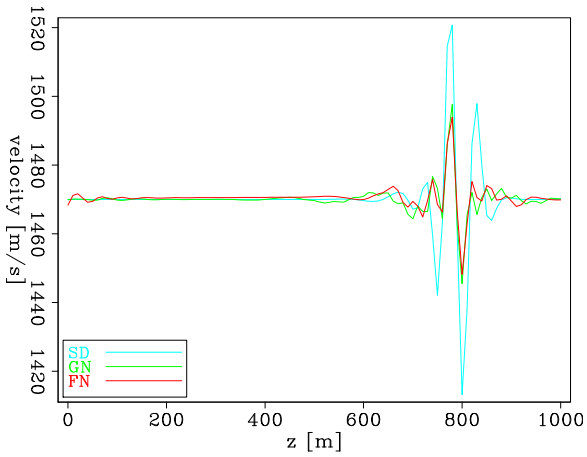


(a)

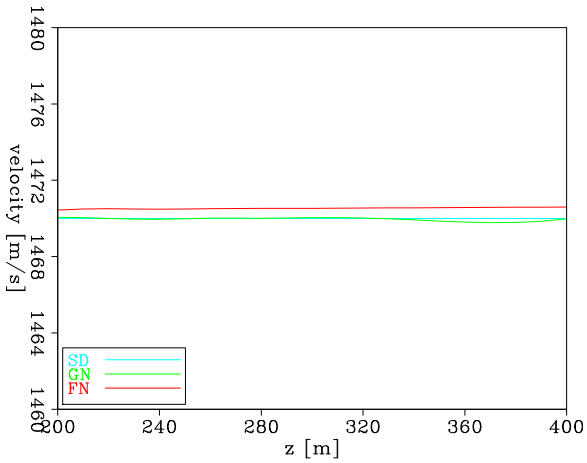
(b)

(c)

Figure 5: (a)Updated model using steepest descent with line search (parabolic fitting). (b) Updated model by inverting the Gauss-Newton Hessian without performing line search. (c) Same as panel (b) but inverting the full Hessian. [**CR**]

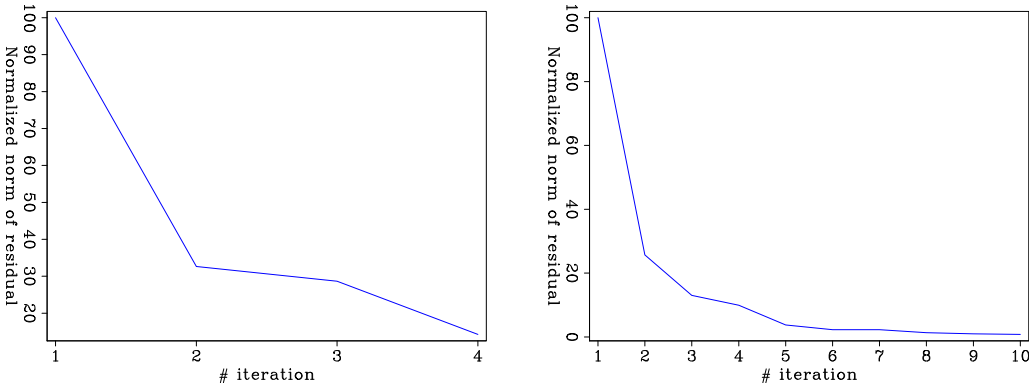


(a)



(b)

Figure 6: (a) Vertical section along middle position in the model for comparing updated model using the three different schemes. (b) Close-up on the depth sections from 200 m to 400 m. Notice that the full-Newton update starts updating also the low-wavenumber component of the model above the reflector. (SD=steepest descent, GN=Gauss-Newton, FN=full-Newton) [**CR**]



(a) (b)

Figure 7: (a) full Hessian iterative inversion convergence, after four iterations the inversion starts diverging. (b) Gauss-Newton Hessian iterative inversion convergence. [**CR**]

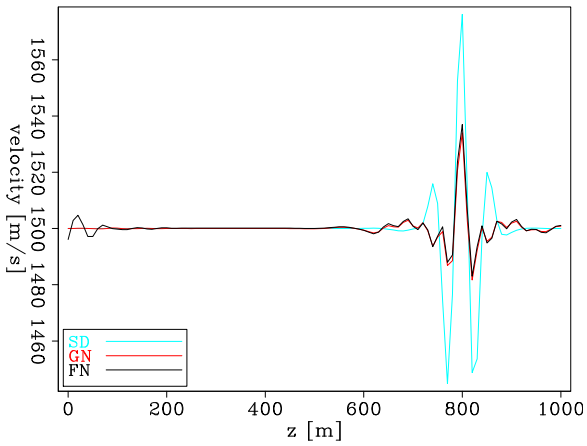


Figure 8: Vertical section along middle position in the model for comparing updated model using steepest descent, the Gauss-Newton, and full-Newton methods using the correct velocity background. As expected Hessian inversions is deconvolving the wavelet effect present in the steepest descent model. It is worth noticing that using the correct background the Gauss-Newton and full-Newton inversion schemes provide almost the same velocity updates. (SD=steepest descent, GN=Gauss-Newton, FN=full-Newton) [**CR**]

Second-iteration results

We computed the gradient and Hessian components for a second iteration of a FWI process. The background model for the second iteration was the one obtained by using steepest descent at the first iteration; that is, the one that is shown in Figure 5(a). Figure 9 illustrates the contributions to the total search direction (opposite of the gradient) of the two components of the data residuals: the modeled and the observed data (equation 24). Figure 9(a) shows the contribution of the modeled data, whereas Figure 9(b) shows the contribution of the observed data. The latter search direction contains a long-wavelength component that correctly tries to increase the background velocity (decrease slowness squared). Figure 9(c) shows the sum of the two components. As for Figure 4, we set the clip to be the same for displaying all the panels in Figure 9 and this clip was chosen to make visible the long-wavelength components of the images.

Figure 10 shows the application of the Gauss-Newton component of the full Hessian to the search direction obtained by computing the total gradient; that is, the search direction shown in Figure 9(c).

Figure 11 shows the application of the four components of the differential WEMVA operator to the search direction shown Figure 9(c): Figure 11(a) shows the $-\mathbf{W}^m\nabla\phi^m$ term (equation 27), Figure 11(b) shows the $+\mathbf{W}^m\nabla\phi^o$ term (equation 28), Figure 11(c) shows the $+\mathbf{W}^o\nabla\phi^m$ term (equation 29), and Figure 11(d) shows the $-\mathbf{W}^o\nabla\phi^o$ term (equation 30). Notice that only the mixed terms (($\mathbf{W}^m\nabla\phi^o$ and $\mathbf{W}^o\nabla\phi^m$) contain significant long-wavelength components; these components have opposite signs between the two panels. Fortunately, the term that has the correct sign ($\mathbf{W}^m\nabla\phi^o$) prevails and the application of the full Hessian (Figure 12) shows a long-wavelength component pointing in the correct direction.

DISCUSSIONS

The theoretical insight presented in the first section of the paper, and illustrated in the last section, has several potential applications. In this section we speculate and elaborate some of these potential applications.

One application could be the improving the convergence of a truncated Newton FWI algorithm when the starting model is inaccurate. As demonstrated in one of the examples above, the WEMVA component of the full Hessian is not immune from cycle skipping when the starting model is sufficiently inaccurate. However, in contrast with the cycle-skipping occurring when the data residuals are backprojected into the gradient, it would be easier to avoid the cycle-skipping of the Hessian. A natural way of avoiding cycle skipping could be to extend the model as it is done in conventional WEMVA and TFWI. Furthermore, the detailed analysis of all the four terms in the WEMVA Hessian (equations 27–30 and Figure 11) could lead to new ways of avoiding the cycle-skipping problem.

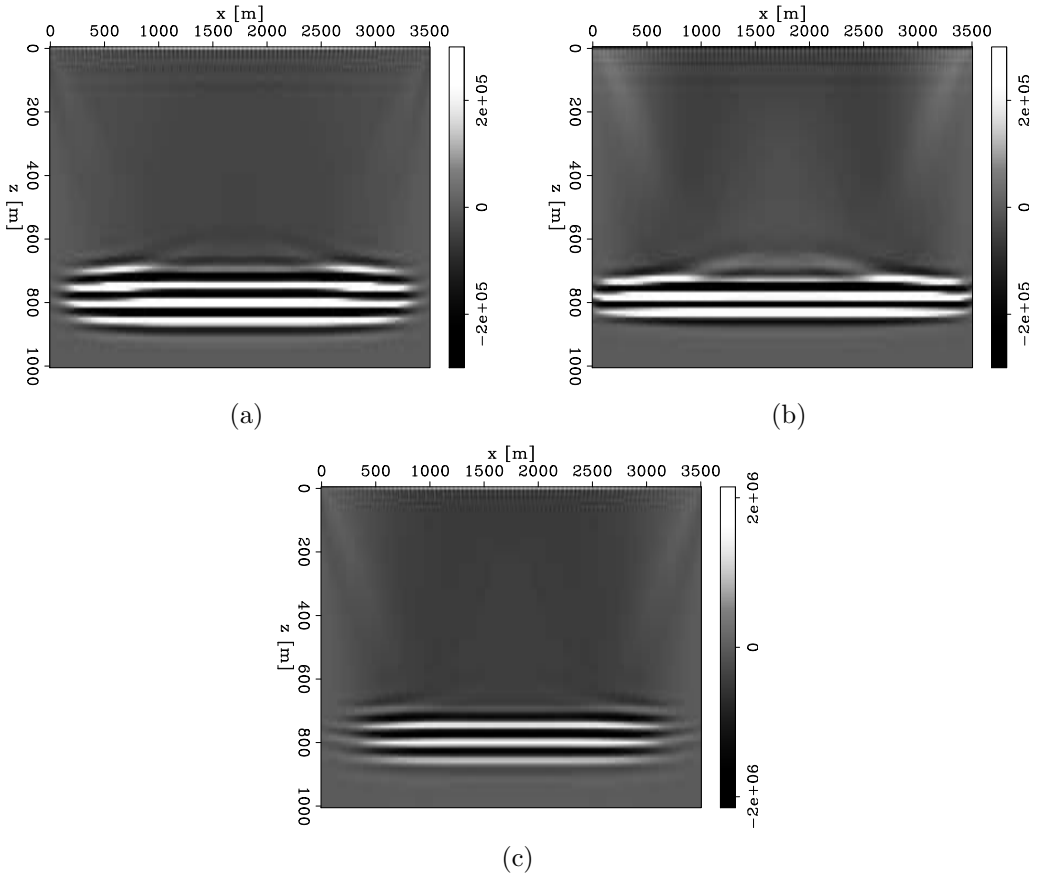


Figure 9: Search direction at the second iteration using the model of Figure 5(a). (a) Search direction given by the modeled data (i.e., $-\nabla\phi^m$). (b) Search direction given by the observed data (i.e., $\nabla\phi^o$). (c) Sum of panel(a) and panel(b) that is the total search direction (i.e., $-\nabla\phi = \nabla\phi^o - \nabla\phi^m$). The long-wavelengths of the total search direction try to decrease the slowness (and increase velocity) on top of the reflector. [**CR**]

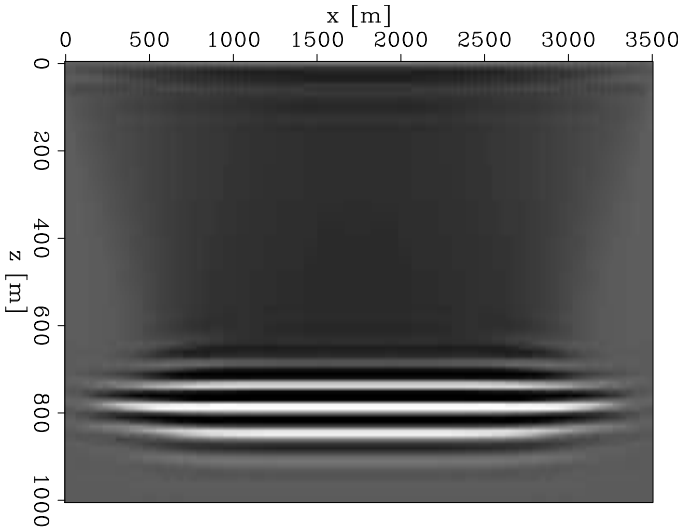


Figure 10: Gauss-Newton Hessian applied to the second search direction (Figure 9(c)). [**CR**]

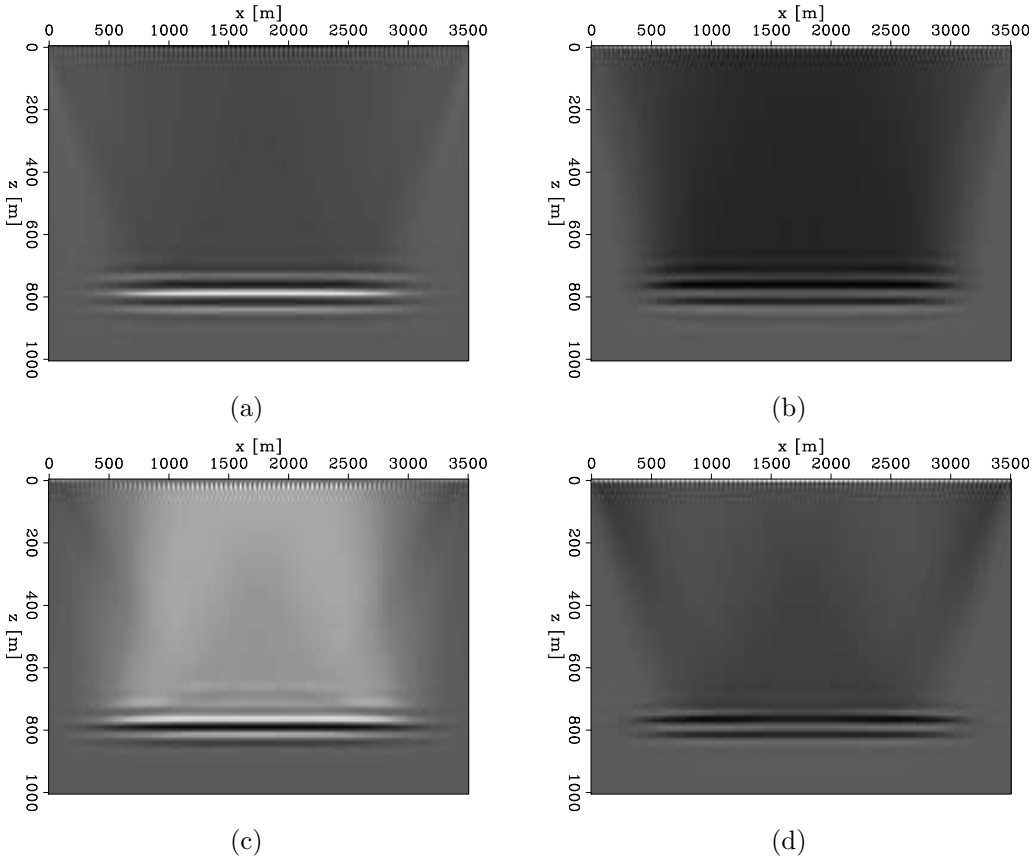


Figure 11: Terms of the WEMVA component of the full Hessian at the second iteration: (a) $-\mathbf{W}^m \nabla \phi^m$ (equation 27). (b) $+\mathbf{W}^m \nabla \phi^o$ (equation 28). (c) $+\mathbf{W}^o \nabla \phi^m$ (equation 29). (d) $-\mathbf{W}^o \nabla \phi^o$ (equation 30). [**CR**]

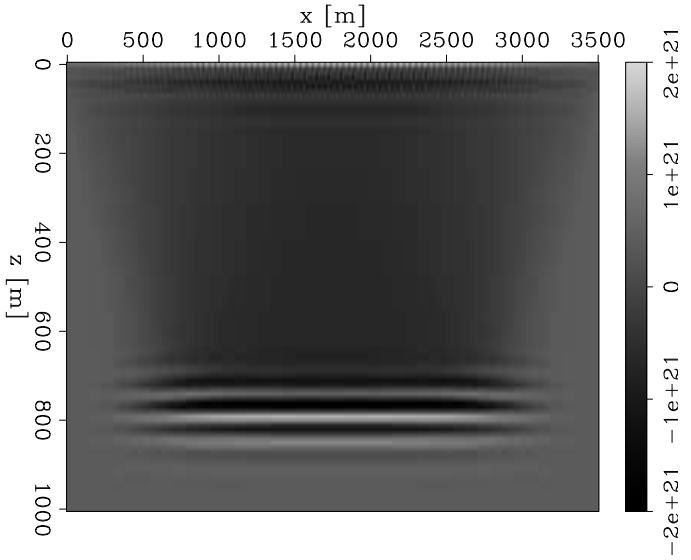


Figure 12: Full Hessian applied to the second search direction (Figure 9(c)). [**CR**]

The decomposition of the Hessian into a Gauss-Newton and a WEMVA component enables the separation of the long wavelengths of the model (velocity) from the short wavelengths (reflectivity) during the inversion. We could subtract the result of applying the inverse of the Gauss-Newton Hessian to the gradient from the application of the whole Hessian, as we did to create the images shown in Figures 4(c) and 4(d). The model updates resulting from this subtraction is dominated by the long-wavelength components of the total update. After the long-wavelengths of the model are accurately estimated, conventional FWI could be safely applied to estimate all the wavelengths in the model.

Another possible application of the insight presented in this paper is to reduce substantially the computational cost of target-oriented fully non-linear elastic FWI solution. Such a solution is attractive because it should lead to a more accurate estimation of elastic reservoir parameters and thus improve reservoir characterization.

If the background model can be assumed to be accurate, the Gauss-Newton component of the Hessian can be precomputed applying one of methods presented by Tang (2011); Deuzeman and Plessix (2015); Fletcher et al. (2012); Tang and Lee (2015), and only the near-range scattering of the differential WEMVA operator needs to be taken into account.

Substantial computational savings can be achieved because repeated applications of the near-range scattering components of the differential WEMVA operator require only local propagation of wavefields. Once the background source wavefield ($\mathbf{u_0}$ in equation 12) is propagated from the source to the target and the data residuals are backpropagated ($\mathbf{u_1^{'}}$ in equation 14) from the surface to the target the application of $\mathbf{K}^{a2}$ and $\mathbf{K}^{b}$ to a model perturbation require only the propagations and scattering of wavefields within the target volume. Therefore, after precomputing the Gauss-

Newton component and the wavefields $\mathbf{u_0}$ and $\mathbf{u_1'}$, the full Hessian can be iteratively inverted only by local propagation and scattering. Although in this paper we discussed only the FWI Hessian related to the acoustic wave equation, this requirement of only local computations holds also in the case of the elastic wave equation. Consequently, expensive long-range elastic propagation are needed only at each non-linear iteration and are not needed for the iterative inversion of the full Hessian. This observation could lead to the development of computationally affordable target-oriented fully-elastic inversion algorithms.

REFERENCES

Alves, G., 2015, Overview of the apache forties data set: SEP-Report, **160**, 173–176.

Ayeni, G., 2011, Time-lapse seismic imaging by linearized joint inversion: PhD thesis, Stanford University.

Baumstein, A., 2014, Extended subspace method for attenuation of crosstalk in multi-parameter full wavefield inversion: SEG Technical Program Expanded Abstracts, 213, 1121–1125.

Biondi, E. and O. O'Reilly, 2015, Two-way wave-equation operators for non-constant density acoustic isotropic media: SEP-Report, **160**, 81–100.

Clapp, M. L., 2005, Imaging under salt: Illumination compensation by regularized inversion: PhD thesis, Stanford University.

Cole, S. and M. Karrenbach, 1992, Least-squares Kirchhoff migration: SEP-Report, **75**, 101–110.

Deuzeman, A. and R.-E. Plessix, 2015, Block-diagonal approximation of the Hessian for multi-parameter FWI: 12th SIAM Conference on Mathematical and Computational Issues in Geosciences, Expanded Abstracts, MS2, Society of Industrial and Applied Mathematics.

Dongarra, J., V. Eijkhout, and A. Kalhan, 1995, Reverse communication interface for linear algebra templates for iterative methods: University of Tennessee, Knoxville, TN.

Epanomeritakis, I., V. Akçelik, O. Ghattas, and B. J., 2008, A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion: Inverse Problems, **24**, 34015–34041.

Fichtner, A., 2011, Full seismic waveform modelling and inversion: Springer Science & Business Media.

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

Fletcher, R. P., S. Archer, D. Nichols, and W. Mao, 2012, Inversion after depth imaging: SEG Technical Program Expanded Abstracts, 546, 1–5.

Jin, S., R. Madariaga, J. Virieux, and G. Lambare, 1991, Two-dimensional asymptotic iterative elastic inversion: SEG Technical Program Expanded Abstracts 1991, 282, 1013–1016.

Kuehl, H. and M. D. Sacchi, 1999, Least-squares split-step migration using the Hartley transform: SEG Technical Program Expanded Abstracts, 396, 1548–1551.

Métivier, L., R. Brossier, J. Virieux, and S. Operto, 2012, The truncated Newton method for full waveform inversion: SEG Technical Program Expanded Abstracts, 480, 1–5.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341–362.

Prucha, M. L., R. G. Clapp, and B. Biondi, 2000, Seismic image regularization in the reflection angle domain: SEP-Report, **103**, 109–119.

Pujol, J., 2007, The solution of nonlinear inverse problems and the levenberg-marquardt method: Geophysics, **72**, no. 4, W1–W16.

Sava, P. and I. Vlad, 2008, Numeric implementation of wave-equation migration velocity analysis operators: Geophysics, **73**, no. 5, VE145–VE159.

Schuster, G. T., 1993, Least-squares cross-well migration: SEG Technical Program Expanded Abstracts, 28, 110–113.

Shen, Y., 2015, Simultaneous inversion of velocity and q using wave-equation migration analysis: SEP-Report, **160**, 71–80.

Tang, Y., 2011, Imaging and velocity analysis by target-oriented wavefield inversion: PhD thesis, Stanford University.

Tang, Y. and S. Lee, 2010, Preconditioning full waveform inversion with phase-encoded Hessian: SEG Technical Program Expanded Abstracts, 204, 1034–1038.

———, 2015, Multi-parameter full wavefield inversion using non-stationary point-spread functions: SEG Technical Program Expanded Abstracts, 215, 1111–1115.

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