Understanding the difference between full Newton and Gauss-Newton approximation of the full waveform inversion Hessian: a short note

Ettore Biondi, Biondo Biondi, and Guillaume Barnier

ABSTRACT

We continue the mixed theoretical-computational study of the difference between full Newton and Gauss-Newton approximation of the Hessian matrix in the context of full waveform inversion (FWI) started in the SEP report 160. We also continue to use an acoustic isotropic wave equation approximation during our discussion. We analyze the connection of the residual-dependent component of the full Hessian with the physical double scattering described by the wave equation. We explain how to avoid inversion instabilities of the full Newton Hessian when this matrix is not positive definitive. With the help of a simple two-perturbation model we study the advantages of full Newton compared to the Gauss-Newton approximation.

INTRODUCTION

Improving convergence rate of any FWI method is fundamental given the high computational cost of the required propagations during the problem optimization (Virieux and Operto, 2009). For this reason many authors have explored the use of the Hessian matrix as a potential preconditioner to reach the problems' optimal solution in the least number of iterations (Pratt et al., 1998; Epanomeritakis et al., 2008; Tang, 2008; Korta et al., 2013; Deuzeman and Plessix, 2015). Therefore, it is critical to understand the connection of each component of this matrix with the physical behavior of any wave equation to find new approximations to be used to precondition FWI optimization.

Newton-like algorithms for non-linear optimization are employed to account for the local curvature of the objective function such that the inverted models present a comparable resolution (i.e., the ability to retrieve a given parameter during the inversion) between model parameters. The differential parameter resolution in the context of multi-parameter FWI is know as inversion crosstalk or trade-off (Operto et al., 2013). Therefore, it is critical to develop an optimal inversion method that utilizes Hessian matrix information.
A detailed computational review of Hessian-based preconditioned FWI has been presented by Métivier et al. (2013). In their work the authors describe different algorithms to perform truncated Newton optimization steps. In their synthetic acoustic

constant-density Marmousi test they find that truncated Gauss-Newton performs better than full Newton approximation. They attribute this difference to slower convergence rate of full Newton matrix during its linear inversion. On the other hand, when applying truncated full Newton to a simpler near-surface FWI problem, the inverted model presents higher spatial resolution when compare to truncated Gauss-Newton using the same number of iterations when inverting the Hessian matrix. From this study, it is therefore unclear whether full Newton has any advantage compare to Gauss-Newton in the case of seismic exploration data inversion. In this report we continue the study shown in Biondi et al. (2015), and review the connection of the full Hessian with the double-scattering mechanism present in any wave equation. We also compare least-square full Newton and Gauss-Newton Hessian matrix inversions for a simple two-perturbation model.

BRIEF ANALYSIS OF THE FWI HESSIAN MATRIX

We start our analysis by defining the classical amplitude-matching FWI objective function as follows:

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

where $\mathbf{m}$ represents the model parameters such as wave-propagation velocities, attenuation, or density, $\mathbf{f}(\mathbf{m})$ is the wave-equation modeling operator, and $\mathbf{d}$ is the observed data. The difference between the latter two is represented by the data residual vector $\mathbf{r}(\mathbf{m})$. It is known that we can decompose our observed data into an infinite scattering series (Weglein et al., 2003). This representation of the data provides an intuitive way to look at the non-linearity present in a typical FWI problem. To do so we can expand the data vector as follows:

$$\mathbf{d} = \mathbf{f}(\mathbf{m}_0) + \frac{\partial \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}}\Delta\mathbf{m} + \frac{1}{2}\Delta\mathbf{m}^* \frac{\partial^2 \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}^2}\Delta\mathbf{m} + \dots, \tag{2}$$

where $\mathbf{m}_0$ is the background model, and $\Delta\mathbf{m}$ is the difference between the true model and the background. This infinite series is effectively a multivariate Taylor expansion of the data vector. The first two terms on the right-hand side of the equation represent the propagation in the background and the first-order scattering (i.e., Born approximation). The additional terms in the series represent the multiple scattering (e.g., multiples and non-linear amplitude behavior).
In the case $\mathbf{m}_0$ is our starting model for a FWI algorithm we see that our initial residuals are given by the terms on the right of the background data $\mathbf{f}(\mathbf{m}_0)$. If the residuals are approximately linear, meaning:

$$\mathbf{r}(\mathbf{m}_0) = \mathbf{d} - \mathbf{f}(\mathbf{m}_0) \approx \frac{\partial \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}}\Delta\mathbf{m}; \tag{3}$$

then, the inverse problem is almost linear, and thus easily solvable up to the null space of the linearized operator. This condition is usually satisfied when the initial model

is reasonably close to the true one, meaning the background model explains most of the observed data in terms of both kinematic and amplitude responses (Biondi and Almomin, 2014).
When a Newton-like scheme is used to minimize equation 1 the following truncated expansion around the initial model is performed:

$$\phi(\mathbf{m}_0 + \Delta\mathbf{m}) \approx \phi(\mathbf{m}_0) + \frac{\partial\phi(\mathbf{m}_0)}{\partial\mathbf{m}}\Delta\mathbf{m} + \frac{1}{2}\Delta\mathbf{m}^*\frac{\partial^2\phi(\mathbf{m}_0)}{\partial\mathbf{m}^2}\Delta\mathbf{m}. \tag{4}$$

The main approximation in the previous relation is to assume a convex quadratic behavior of the objective function close to the initial model. We already see that if this assumption is not fulfilled, any Newton method may fail completely to find the local minimum of the objective function. The key for the success of this class of algorithms resides in the Hessian matrix of the objective function $\frac{\partial^2\phi(\mathbf{m}_0)}{\partial\mathbf{m}^2}$ being positive definitive. In fact, in any Newton-like optimization we try to solve the following linear system:

$$\frac{\partial^2\phi(\mathbf{m}_0)}{\partial\mathbf{m}^2}\Delta\mathbf{m} = -\frac{\partial\phi(\mathbf{m}_0)}{\partial\mathbf{m}}. \tag{5}$$

This system is obtained by setting the derivative of equation 4 with respect to the model perturbation $\Delta\mathbf{m}$ to zero. It is known that the full Hessian matrix can be decomposed into its Gauss-Newton approximation plus an additional residual-dependent component (Fichtner, 2010). In fact, we observed that:

$$\frac{\partial^2\phi(\mathbf{m})}{\partial\mathbf{m}^2} = \left(\frac{\partial\mathbf{f}(\mathbf{m})}{\partial\mathbf{m}}\right)^*\left(\frac{\partial\mathbf{f}(\mathbf{m})}{\partial\mathbf{m}}\right) + \left(\frac{\partial^2\mathbf{f}(\mathbf{m})}{\partial\mathbf{m}^2}\right)^*\mathbf{r}(\mathbf{m}), \tag{6}$$

where the first term is composed of the product between forward and adjoint Born operator, and the second one is connected to the second-order scattering in the series of equation 2. Because of this connection many authors see an additional value to the full Hessian when a Newton-like method is employed in solving FWI problems (Pratt et al., 1998; Epanomeritakis et al., 2008). It is still under debate whether this additional value is meaningful or not in seismic exploration (Métivier et al., 2013). In addition, as shown by Biondi et al. (2015) the residual-dependent component of the Hessian matrix is connected to the wave-equation migration velocity analysis (WEMVA) operator (Sava and Vlad, 2008).

We now discuss a few considerations when a single full-Newton step is performed during full waveform inversion. The Newton system of equations at the first iteration is given by:

$$\left[\left(\frac{\partial\mathbf{f}(\mathbf{m}_0)}{\partial\mathbf{m}}\right)^*\left(\frac{\partial\mathbf{f}(\mathbf{m}_0)}{\partial\mathbf{m}}\right) + \left(\frac{\partial^2\mathbf{f}(\mathbf{m}_0)}{\partial\mathbf{m}^2}\right)^*\mathbf{r}(\mathbf{m}_0)\right]\Delta\tilde{\mathbf{m}} = -\left(\frac{\partial\mathbf{f}(\mathbf{m}_0)}{\partial\mathbf{m}}\right)^*\mathbf{r}(\mathbf{m}_0). \tag{7}$$

The ideal solution of this linear inversion would be:

$$\Delta\tilde{\mathbf{m}} = \Delta\mathbf{m}, \tag{8}$$

meaning we converge to the true perturbation of the data expansion around the initial FWI model (equation 2) in one Newton step. Unless our problem is close to being

linear this result is difficult to find. Assuming that most of the non-linearities of an FWI problem are given by the second-order scattering, we can truncate the Taylor expansion to the first three terms and make the inversion closer to linearity. The initial residuals are thus given by the following:

$$\mathbf{r}(\mathbf{m}_0) = -\frac{\partial \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}}\Delta\mathbf{m} - \frac{1}{2}\Delta\mathbf{m}^*\frac{\partial^2 \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}^2}\Delta\mathbf{m}. \tag{9}$$

Substituting equation 9 into 7 we obtain:

$$\begin{aligned}&\left[\left(\frac{\partial^2 \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}^2}\right)^*\left(-\frac{\partial \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}}\Delta\mathbf{m} - \frac{1}{2}\Delta\mathbf{m}^*\frac{\partial^2 \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}^2}\Delta\mathbf{m}\right)\right]\Delta\tilde{\mathbf{m}} \\ &= \frac{1}{2}\left(\frac{\partial \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}}\right)^*\Delta\mathbf{m}^*\frac{\partial^2 \mathbf{f}(\mathbf{m}_0)}{\partial \mathbf{m}^2}\Delta\mathbf{m},\end{aligned} \tag{10}$$

where we notice that the right-hand side term is basically the image produced by the first-order multiples in the data. The term on the left is more complicated, but can be decomposed into two different operators; a WEMVA-like operator applied to the true Born data and the second-order scattering, respectively. We will converge in a single Newton step to the correct solution if equality 8 holds. In the next section we explore these considerations on a simple two-perturbation synthetic test.

SYNTHETIC TEST

In this section, we show an example of FWI in which a single truncated Newton step is performed on a simple synthetic model where two perturbations of 100 m/s are added to a constant velocity background of 2000 m/s (Figure 1). We inject a Ricker wavelet with central frequency of 20 Hz and use evenly spaced receivers and sources at the surface by 10 m and 100 m, respectively. A shot gather containing reflection only data is shown in Figure 2a. The choice of this model introduces a mild non-linearity when the correct background velocity is considered as starting model for the inversion. In fact, comparing the observed data with the correct Born approximation (i.e., second term in the expansion 2) we clearly see that most of the recorded amplitudes are represented by this linear modeling operator (Figure 2b). The comparison of the difference between non-linear data and Born approximation with the second-order scattering term in the series provides us with an insight of the non-linearities in the treated FWI problem. From Figure 2c we observe that most of the higher-order scattering is contained in the second-order one. The amplitudes of these events are approximately one order of magnitude smaller than the first-order approximation (compare Figures 2b and 2c). Because of these observations we decide to consider only first- and second-order scattering in this test.

First of all we compute the initial search direction given the recorded data. Figure 3a shows the search direction obtained using both orders of scattering. The maximum of this search direction corresponds to the peaks of the used perturbations. To understand the effect of the presence of second-order scattering in the data we

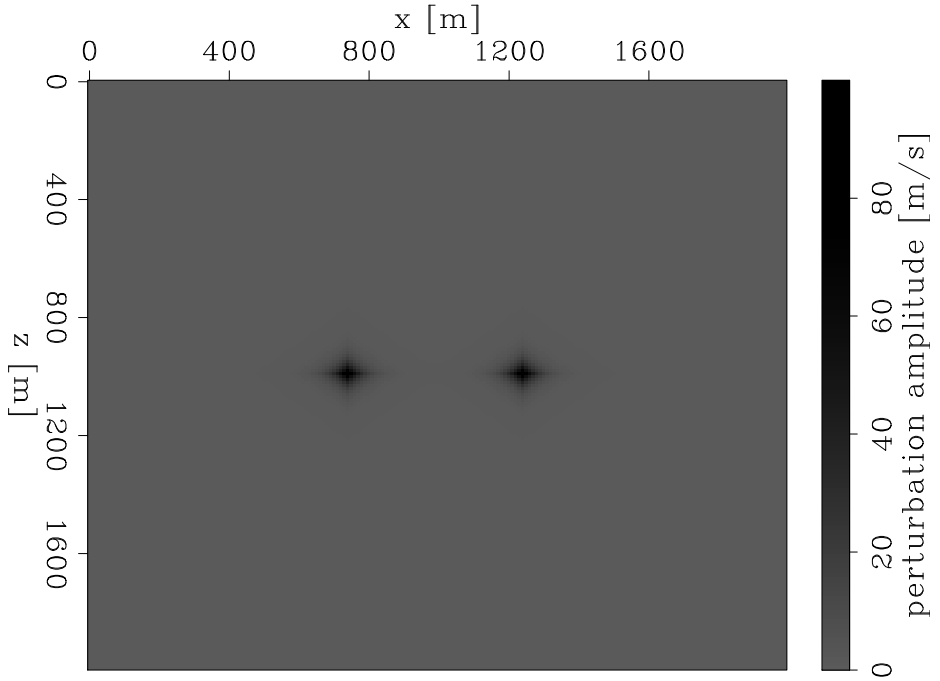

Figure 1: True velocity perturbation used during the reported synthetic test. The background model is a constant velocity of 2000 m/s. [**ER**]

compute the image of this energy (Figure 3b). We first see that the relative amplitude is approximately 90 percent weaker compared to the total image. In addition, the phase of the imaged points is slightly different than the one in the total search direction. This plot describes the image mainly produced by the self-interaction of the scattered wavefield with the perturbation. Because of the chosen perturbation, the image of the interaction between the two anomalies is almost negligible.

As explained in the previous section, the hope of converging to the original model perturbation in a single Newton step depends on matching the image of the multiples by the application of the residual-dependent Hessian component to the true perturbation (equation 10). Figure 4a displays this application. Despite the similar amplitude behavior, the phase of the image is quite different; thus, it is unlikely that a single Newton step will provide the correct answer even though the mild non-linearity of the problem. Again, the contribution of the second-order scattering in the application of this Hessian component is weaker with respect to the first-order scattering residuals (compare Figures 4a and 4b).

We test two different approximations of the FWI Hessian matrix against the full Newton one. We compare Gauss-Newton, full Newton, and a Hessian approximation obtained by a matrix expansion. In fact, by decomposing the full Hessian $\mathbf{H}$ into its Gauss-Newton $\mathbf{H}_{GN}$ and residual-dependent $\mathbf{H}_r$ components, we can write the following inverse approximation:

$$(\mathbf{H})^{-1} = (\mathbf{H}_{GN} + \mathbf{H}_r)^{-1} = \left(\mathbf{I} + \mathbf{H}_{GN}^{-1}\mathbf{H}_r\right)^{-1} \mathbf{H}_{GN}^{-1} \approx \left(\mathbf{I} - \mathbf{H}_{GN}^{-1}\mathbf{H}_r\right) \mathbf{H}_{GN}^{-1}, \qquad (11)$$

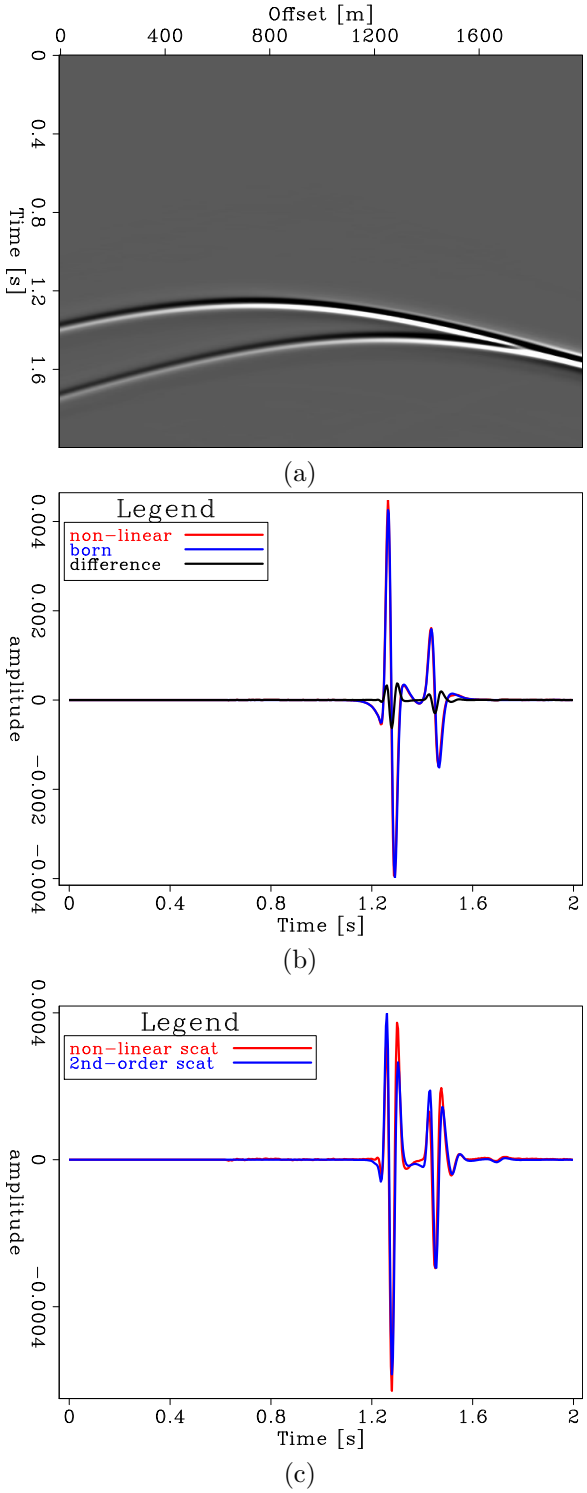

Figure 2: (a) Reflection data observed for one shot when the velocity perturbation in Figure 1 is used. (b) Comparison between non-linear data (red curve) and Born approximation (blue curve) using the true perturbation for a receiver at 1000 m. The black curve represents the higher-order scattering effects present in the observed data. (c) Comparison between true second-order scattering and black curve in Figure 2b. The direct arrival has been removed because we assume to start the inversion with the correct background. [**ER**]

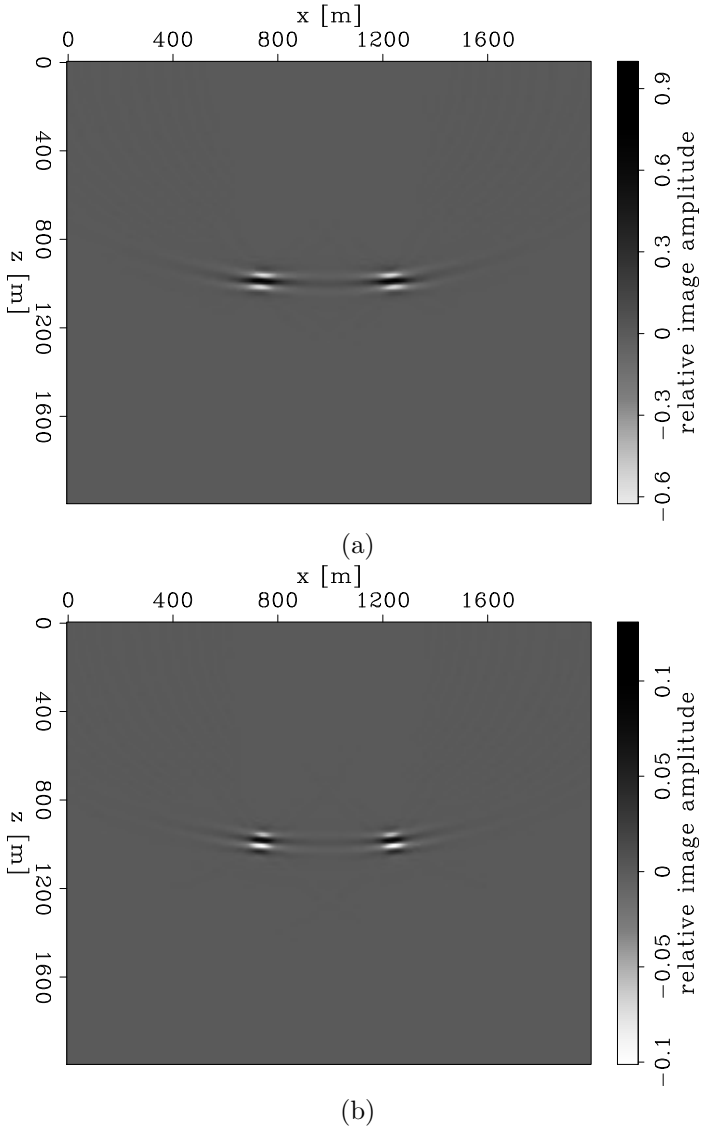

Figure 3: (a) First search direction when only first- and second-order scattering is considered to be present in the recorded data. (b) First search direction due to the second-order scattering in the data (i.e., right-hand side term in equation 10). [**ER**]

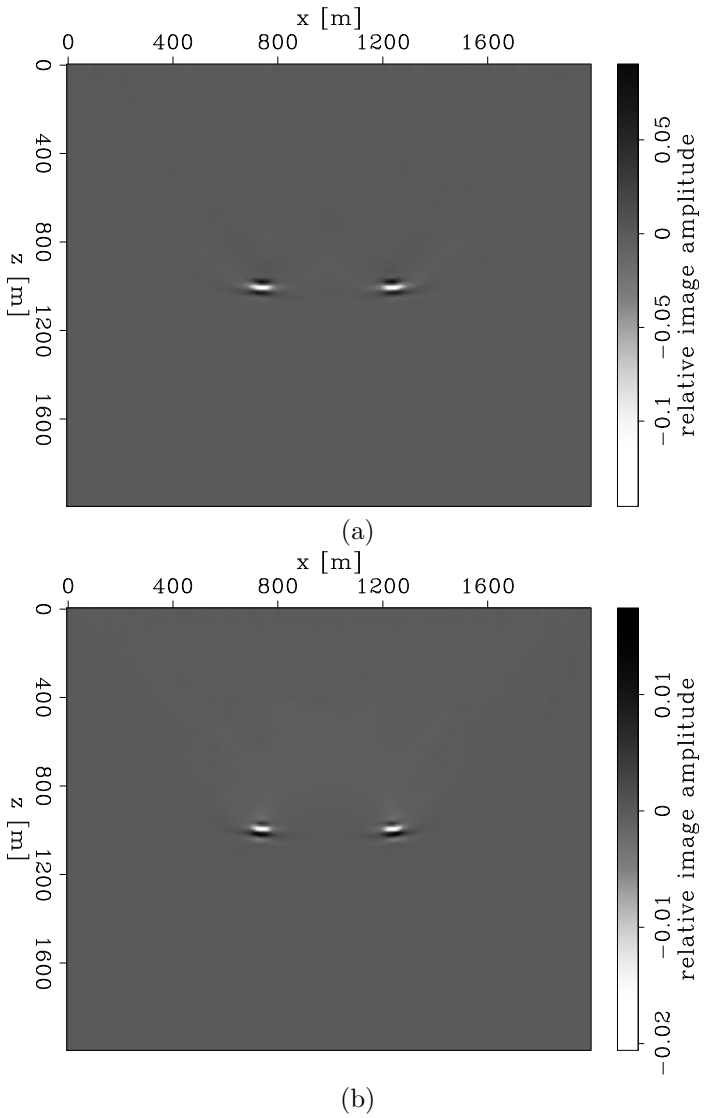

Figure 4: Application of the residual-dependent hessian component to the true perturbation using the total residuals (a) and the second-order scattering component (b). The amplitudes are normalized with respect to the image shown in Figure 3a [**ER**]

where we assume that the inverse for the Gauss Newton component exists, and that the matrix norm $\|\mathbf{H}_{GN}^{-1}\mathbf{H}_r\|$ is negligible compared to the norm of an identity matrix with the same size. From this equation we see that a correction factor, which is given by the product of a WEMVA-like operator and Gauss-Newton inverse matrix, is subtracted from the inverted perturbation.

In all of these linear inversions the different matrices are solved iteratively in the least-squares sense as follows:

$$\phi(\Delta\mathbf{m}) = \frac{1}{2}\left\|\mathbf{H}(\mathbf{m}_0)\Delta\mathbf{m} - \Delta\mathbf{m}_{mig}(\mathbf{m}_0)\right\|_2^2, \tag{12}$$

where $\mathbf{H}$ is the Hessian matrix to be inverted, and $\Delta\mathbf{m}_{mig}(\mathbf{m}_0)$ is the first FWI search direction (i.e., Figure 3a in this test). We choose this approach to mitigate the problem of the full Hessian matrix being symmetric indefinite (i.e., having positive and negative eigenvalues); an issue that has been also observed by Métivier et al. (2013). In fact, when we employ the symmetric solver proposed by Biondi and Barnier (2017) on the full Hessian matrix, after few iterations, the inversion starts diverging and becoming unstable. This least-squares approach stabilizes the inversion, although with this approach we are squaring the condition number of the actual inverted matrix.

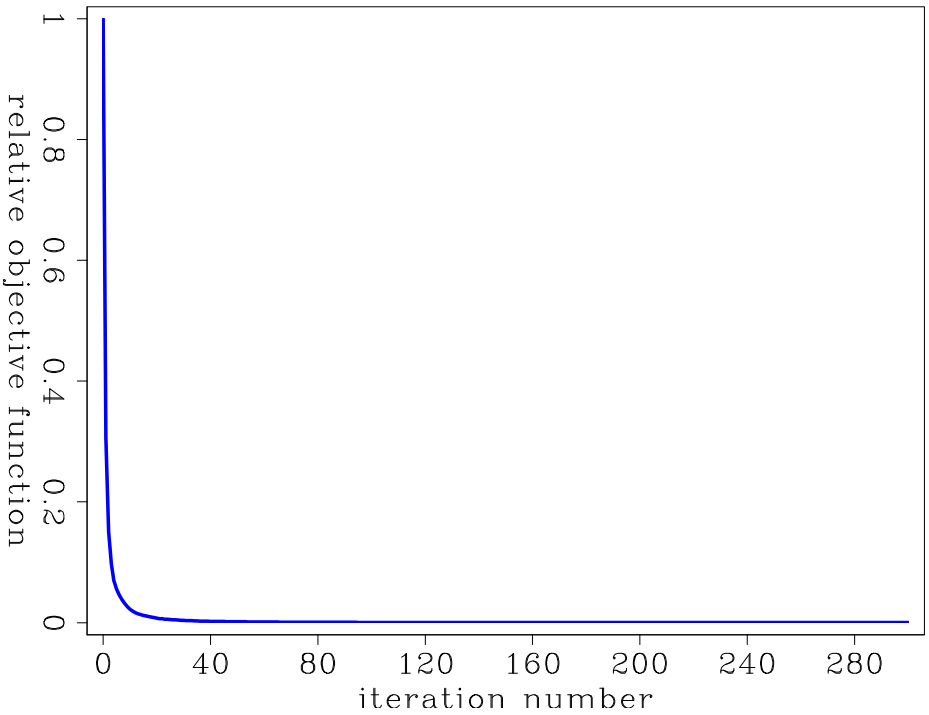

Figure 5: Relative objective function value for least-squares inversion of Gauss-Newton Hessian matrix. A similar behavior is seen for all the other described truncated Hessian inversions. [**CR**]

In the first Hessian inversion we consider a Gauss-Newton approximation. Figure 5 shows the relative variation of the objective function 12 when 300 iterations of

conjugate-gradient algorithm are performed. At the last iteration the objective function value is five orders of magnitude smaller than the initial one. Figure 6a shows the inverted perturbation at the last iteration. First, we observe that the amplitude of the anomalies is underestimated. Secondly, despite the considerably decrease in the objective function, the wavelet signature is still present in the inverted anomalies. When this inverted perturbation is reprojected non-linearly in the data space, the modeled amplitudes closely match the recorded ones (Figure 6b), meaning that most of the linear component of the data has been inverted and potentially we are encountering a null space of the linear operator. We then inverted iteratively the full Newton hessian and the matrix shown in equation 11. Figure 6c shows the comparison of these inversions on a depth profile passing through the left-hand anomaly. From this comparison we observe that practically there is no difference between full Newton and Gauss-Newton inversion results. The similar behavior of the solutions could be attributed to the values of the Gauss-Newton matrix elements that dominate the inversion. On the other hand, we the full Newton approximation shown in equation 11 provides a model slightly closer to the true perturbation. This fact potentially means that the correction term in this approximation could enhance the additional value of the full Newton matrix in the residual-dependent component and avoid the instabilities possibly present in the full Hessian. In fact, in this case we are inverting the more stable positive semi-definitive Gauss-Newton matrix (see equation 11).

CONCLUSIONS AND FUTURE DIRECTIONS

We discuss the difference between the full Hessian and its Gauss-Newton approximation in the context of FWI. We show the connection of the residual-component of this matrix to the second-order scattering present in the data. We also discuss the value given by truncated Newton steps and when those could potentially converge to the true perturbation in one single matrix inversion. With the help of a simple two-anomaly model we study the effect of a single Hessian matrix inversion step. Moreover, with the proposed approximation of the full Newton matrix we are able to enhance the residual-dependent component and avoid unstable inversion results. In the future, we are going to study the advantage of performing truncated Newton steps during a non-linear inversion framework. In addition, further studies of the proposed full Hessian approximation are going to be made to better evaluate its potential in more complex inversion scenarios.

REFERENCES

Biondi, B. and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, WA129–WA140.

Biondi, B., E. Biondi, M. Maharramov, and Y. Ma, 2015, Dissection of the full-waveform inversion Hessian: SEP-Report, **160**, 81–100.

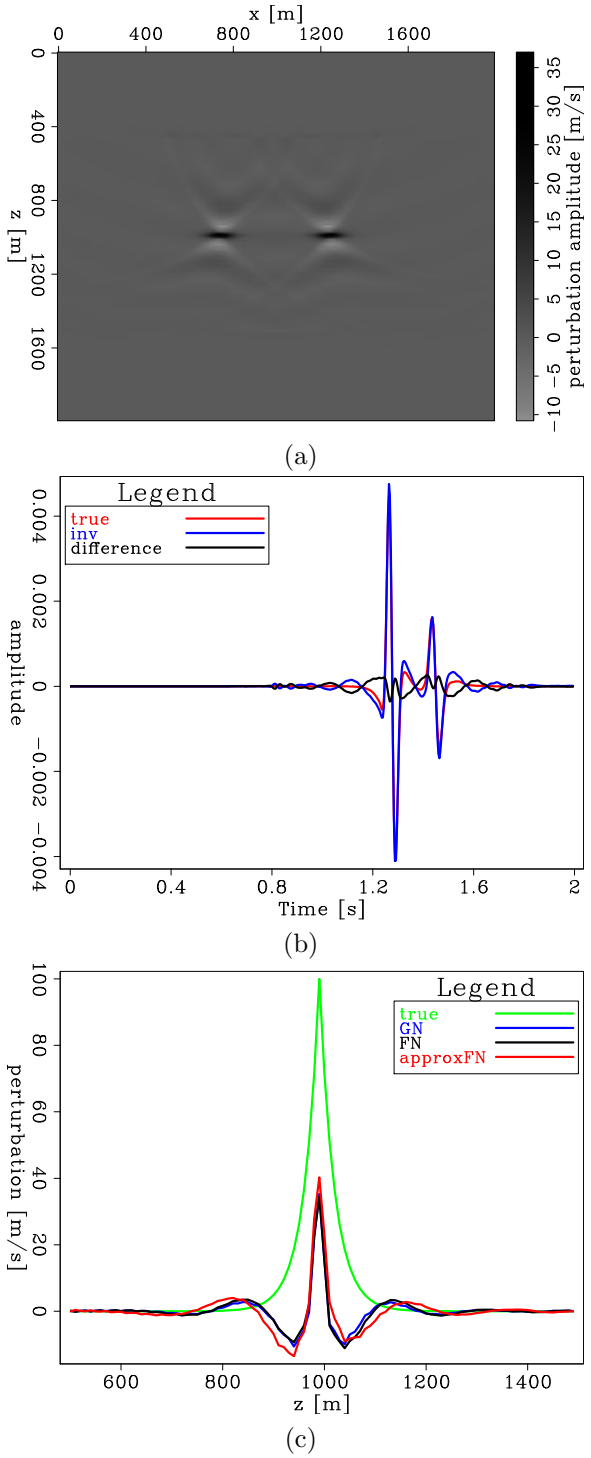

Figure 6: (a) Inverted perturbations using a Gauss-Newton approximated Hessian. (b) Comparison between observed data (red curve), data generated using the inverted perturbations (blue curve) in Figure 6a, and their difference (black curve). (c) Comparison between Gauss-Newton (blue curve), full Newton (black curve), and approximated full Newton (red curve) of the inverted perturbations against the true anomalies (green curve). [**CR**]

Biondi, E. and G. Barnier, 2017, A flexible out-of-core solver for linear/non-linear problems: SEP-Report, **168**.

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.

Epanomeritakis, I., V. Akçelik, O. Ghattas, and J. Bielak, 2008, A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion: Inverse Problems, **24**, 034015.

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

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.

Métivier, L., R. Brossier, J. Virieux, and S. Operto, 2013, Full waveform inversion and the truncated Newton method: SIAM Journal on Scientific Computing, **35**, B401–B437.

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.

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

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

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

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Weglein, A. B., F. V. Araújo, P. M. Carvalho, R. H. Stolt, K. H. Matson, R. T. Coates, D. Corrigan, D. J. Foster, S. A. Shaw, and H. Zhang, 2003, Inverse scattering series and seismic exploration: Inverse problems, **19**, R27.