Approximate- vs. full-Hessian in FWI: 1D analytical and numerical experiments

Raul Cova and Kris Innanen

ABSTRACT

Feasibility of using Full Waveform Inversion (FWI) to build velocity models has been increasing as the computational power and more comprehensive forward modelling approaches have arose. Traditionally, the objective function used during the FWI process has been minimized by the means of steepest-descent methods by conditioning the gradient function. Quasi- and full-Newton methods use the Hessian, or an approximate to it, as a gradient conditioning. For a 1D scalar medium, we derived an analytical expression for the approximate-Hessian suggesting that it brings the model update to within a first order approximation of the exact reflection coefficient for a single interface. In its functional form the Hessian is represented by delta functions into the integration for computing the model update. Compared to the approximate-Hessian, we found that the full-Hessian provides additional scaling information at the depth of the interface, improving the accuracy of the inversion. These ideas were also tested using a numerical example displaying how both Hessians move very fast toward the actual velocity model. It is also shown that the full-Hessian leads to a very accurate inversion in the presence of large velocity contrasts superior to the approximate-Hessian. Hence, the full-Hessian may achieves a faster convergence and accurate inversion while providing amplitude information. For large velocity contrasts, or in a 2D case, where strong AVO effects may be present, the application of the full-Newton FWI might be a good candidate.

INTRODUCTION

Approaches to Full Waveform Inversion (FWI) have evolved since the initial works of Tarantola (1984) and Lailly (1983) using the steepest-descent method for minimizing the objective function in the inversion of seismic waveforms. Virieux and Operto (2009) compiled a very comprehensive overview of the theory and challenges behind FWI. The role of the Hessian in the FWI algorithms is one of these challenges. Pratt et al. (1998) derived a matrix formalism for FWI regarding the use of the full-Hessian (Newton method) or some approximation to the Hessian (Quasi-Newton method). Margrave et al. (2011) developed a framework for interpreting the influence of the Hessian into FWI in a functional form in terms of Green's functions.

In this paper, for a 1D scalar medium, we show an analytical derivation of the Hessian and its inverse. Using numerical examples we also show, the effect of the approximate- and full-Hessian into the FWI process.

Minimizing the objective function in FWI

Under scalar medium assumptions, the goal of FWI is to find the squared-slowness model $s_0(\mathbf{r}) \equiv 1/c_{0,n}^2(\mathbf{r})$ which "most likely" produced, the recorded wavefield $P(\mathbf{r}_g, \mathbf{r}_s, \omega)$. Here, $\mathbf{r}_g$ and $\mathbf{r}_s$ are the receiver and source position vectors respectively, and ω is the angu-

lar frequency. In order to achieve that goal, FWI seeks to minimize, after n iterations, the objective function $\Phi(s_0^{(n)})$ defined as

$$\Phi(s_0^{(n)}) \equiv \frac{1}{2}\int d\omega \left(\sum_{s,g} |\delta P|^2\right), \tag{1}$$

where δP is the residual wavefield given by

$$\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)}) \equiv P(\mathbf{r}_g, \mathbf{r}_s, \omega) - G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)}), \tag{2}$$

and $G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)})$ is the modelled field due to the $s_0^{(n)}$ squared-slowness model iteration.

After evaluating the objective function at $\Phi(s_0^{(n)} + \delta s_0^{(n)})$, computing its derivative and setting it to zero, Margrave et al. (2011), showed that the update $\delta s_0^{(n)}$ needed to move toward the minimum of the objective function can be computed as

$$\delta s_0^{(n)}(r'') = -\int d\mathbf{r}' H^{(n)-}(\mathbf{r}'', \mathbf{r}') g^{(n)}(\mathbf{r}'), \tag{3}$$

where

$$g^{(n)}(\mathbf{r}') = \frac{\partial \Phi(s_0^{(n)})}{\partial s_0^{(n)}(\mathbf{r}')}, \tag{4}$$

is the gradient of the objective function and $H^{(n)-}(\mathbf{r}'', \mathbf{r}')$ is the inverse of the Hessian function defined as

$$H^{(n)}(\mathbf{r}', \mathbf{r}) = \frac{\partial^2 \Phi(s_0^{(n)})}{\partial s_0^{(n)}(\mathbf{r}') \partial s_0^{(n)}(\mathbf{r})}, \tag{5}$$

$\mathbf{r}, \mathbf{r}'$ and $\mathbf{r}''$, are arbitrary positions into the slowness model.

Margrave et al. (2011) also showed that, in terms of Green's functions $G(\mathbf{r}, \mathbf{r}', \omega)$, the Gradient can be computed as

$$g^{(n)}(\mathbf{r}) = \sum_{s,g} \int d\omega \omega^2 [G(\mathbf{r}, \mathbf{r}_s, \omega | s_0^{(n)})] \times [G(\mathbf{r}_g, \mathbf{r}, \omega | s_0^{(n)}) \delta P^*(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)})], \tag{6}$$

while the Hessian can be expressed as composed of two terms,

$$H^{(n)}(\mathbf{r}', \mathbf{r}) = H_1^{(n)}(\mathbf{r}', \mathbf{r}) + H_2^{(n)}(\mathbf{r}', \mathbf{r}), \tag{7}$$

where

$$\begin{aligned} H_1^{(n)}(\mathbf{r}', \mathbf{r}) = -\sum_{s,g} \int d\omega \omega^4 [&G(\mathbf{r}_g, \mathbf{r}, \omega) G(\mathbf{r}, \mathbf{r}', \omega) G(\mathbf{r}', \mathbf{r}_s, \omega) \\ &+ G(\mathbf{r}_g, \mathbf{r}', \omega) G(\mathbf{r}', \mathbf{r}, \omega) G(\mathbf{r}, \mathbf{r}_s, \omega)] \delta P^*(\mathbf{r}_g, \mathbf{r}_s, \omega)], \end{aligned} \tag{8}$$

and

$$H_2^{(n)}(\mathbf{r}', \mathbf{r}) = \sum_{s,g} \int d\omega \omega^4 G(\mathbf{r}_g, \mathbf{r}', \omega) G(\mathbf{r}', \mathbf{r}_s, \omega) G^*(\mathbf{r}_g, \mathbf{r}, \omega) G^*(\mathbf{r}, \mathbf{r}_s, \omega)). \tag{9}$$

Equation 7 is known as the full-Hessian and it is used for taking full-Newton steps towards the minimum of the objective function. Under the assumption of small residuals, equation 8 can be small enough to be omitted in the computations, leading to an approximate-Hessian equal to equation 9. The latter one is used for taking quasi-Newton steps for FWI. Dependency of the fields and residuals on $s_0^{(n)}$ is implicit in equations 8 and 9 and will be omitted in the following equation to save space.

1D Approximate-Hessian in a scalar medium

The approximate Hessian (equation 9) can be written in terms of Green's functions and its complex conjugates, representing the receivers and source wavefields observed at positions $\mathbf{r}$ and $\mathbf{r}'$. The general expression for a causal homogeneous Green's function, in a 1D scalar medium is

$$G(z, z_s, \omega) = \frac{e^{ik|z-z_s|}}{i2k}, \tag{10}$$

where z and z_s are the depths of the observation point and source point respectively, and k is the wave number in the direction of propagation.

Figure 1 shows the interface model used for deriving the analytical expressions for the Hessian; (left) shows a representation of the background homogeneous medium with velocity c_0 and (right) shows the true model with just one interface at depth z_1.

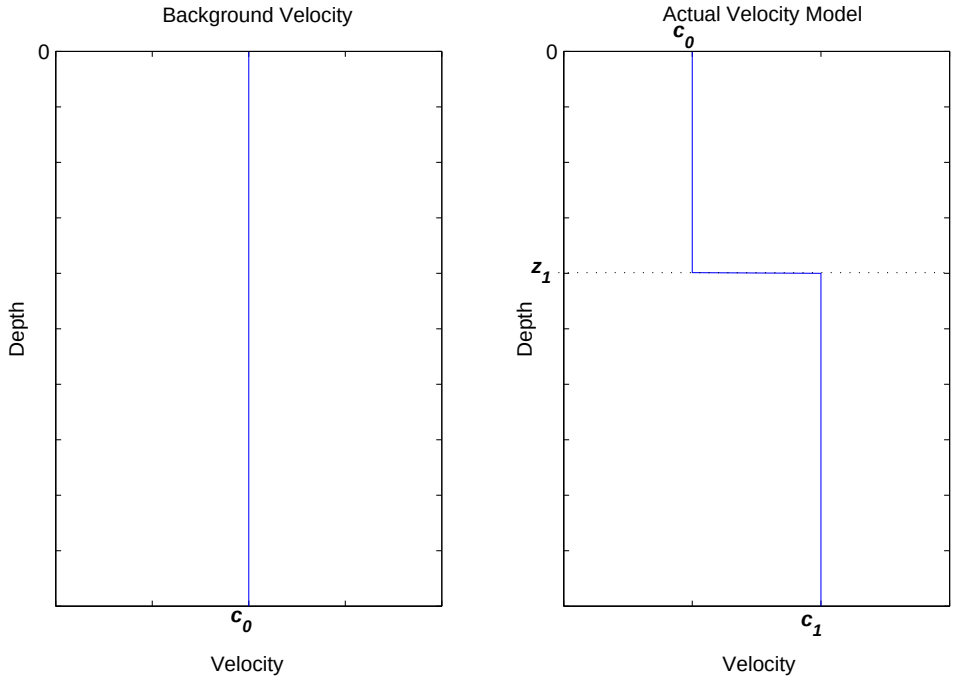

FIG. 1. (left) Background velocity model used as the initial model for 1D FWI. (right) Actual velocity model we are trying to invert

For the initial conditions mentioned above and considering just one source-receiver pair fixed at depth $z_g = z_s = 0$, the approximate-Hessian can be written as

$$H_2(z', z) = \int d\omega \omega^4 G(0, z', \omega) G(z', 0, \omega) G^*(0, z, \omega) G^*(z, 0, \omega)). \tag{11}$$

Using equation 10, the two modelling Green's functions and their complex conjugate, for a wavefield propagating through an background medium with wave number $k = k_0$ are

$$G(z, 0, \omega) = G(0, z, \omega) = \frac{e^{ik_0 z}}{i2k_0}, \tag{12}$$

$$G^*(z, 0, \omega) = G^*(0, z, \omega) = -\frac{e^{-ik_0 z}}{i2k_0}. \tag{13}$$

Substituting equations 12 and 13 into equation 11 we get

$$H_2(z', z) = \int d\omega \omega^4 \frac{e^{ik_0 z'}}{i2k_0} \frac{e^{ik_0 z'}}{i2k_0} \left(-\frac{e^{-ik_0 z}}{i2k_0}\right) \left(-\frac{e^{-ik_0 z}}{i2k_0}\right). \tag{14}$$

Multiplying and collecting alike terms, equation 14 reads into

$$H_2(z', z) = \frac{1}{2^4} \int d\omega \left(\frac{\omega}{k_0}\right)^4 e^{i2k_0(z'-z)}. \tag{15}$$

Since $w/k_0 = c_0$ we can take this term out of the integral and using the variable change

$$d\omega = \frac{c_0}{2} d(2k_0), \tag{16}$$

equation 15 becomes

$$H_2(z', z) = \left(\frac{c_0}{2}\right)^5 \int d(2k_0) e^{i(2k_0)(z'-z)}. \tag{17}$$

The remaining integral over the variable $(2k_0)$ can be recognized as the Fourier transform of a delta function at point $(z' - z)$, then the approximate-Hessian becomes

$$\begin{aligned} H_2(z', z) &= \left(\frac{c_0}{2}\right)^5 (2\pi)\delta(z' - z), \\ &= \frac{c_0^5 \pi}{2^4} \delta(z' - z). \end{aligned} \tag{18}$$

Inverse Approximate 1D Hessian

The inverse Hessian must satisfy the expression (Margrave et al., 2011)

$$\int d\mathbf{r}' H^{(n)}(\mathbf{r}'', \mathbf{r}') H^{(n)-}(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r}'' - \mathbf{r}). \tag{19}$$

Rewriting equation 19 for the 1D case, using equation 18 gives

$$\int dz' H(z'', z') H^-(z, z') = \delta(z'' - z), \tag{20}$$

$$\int dz' \left[\frac{c_0^5 \pi}{2^4} \delta(z'' - z')\right] H^-(z, z') = \delta(z'' - z), \tag{21}$$

$$\frac{c_0^5 \pi}{2^4} \int dz' \delta(z'' - z') H^-(z, z') = \delta(z'' - z). \tag{22}$$

Using the sifting property of the delta function,

$$\frac{c_0^5\pi}{2^4}H^-(z,z'') = \delta(z''-z), \tag{23}$$

and solving for $H^-(z,z'')$,

$$H^-(z,z'') = \frac{2^4}{c_0^5\pi}\delta(z''-z). \tag{24}$$

Model update given by the approximate-Hessian

In order to fully understand the effect of the Hessian in updating the initial model, the action of it on the gradient (equation 3) should be known.

The gradient of the objective function $\Phi(s_0^{(n)})$ for a 1D scalar homogeneous medium is given by (Innanen, 2013)

$$g(z) = \frac{R_1 c_0^3\pi}{4}\mathrm{H}(z-z_1), \tag{25}$$

where R_1 is the scalar reflection coefficient for normal incidence computed as $R_1 = (c_1-c_0)/(c_1+c_0)$, and $\mathrm{H}(z-z_1)$ is a Heaviside function with a step located at the reflector depth z_1.

Rewriting equation 3 for the 1D scenario and substituting equations 24 and 25 into it

$$\delta s_0^{(n)}(z) = -\int dz' H^-(z,z')g(z'), \tag{26}$$

$$= -\int dz' \left[\frac{2^4}{c_0^5\pi}\delta(z'-z)\right]\left[\frac{R_1 c_0^3\pi}{4}\mathrm{H}(z'-z_1)\right], \tag{27}$$

$$= -\frac{4R_1}{c_0^2}\int dz'\delta(z'-z)\mathrm{H}(z'-z_1). \tag{28}$$

invoking the sifting property of the delta function

$$\delta s_0^{(n)}(z) = -\frac{4R_1}{c_0^2}\mathrm{H}(z-z_1). \tag{29}$$

Equation 29 shows that the model update after using the inverse Hessian acting over the gradient of the objective function, gives a step with amplitude $-4R_1/c_0^2$ at the actual depth of the interface z_1.

In order to understand the significance of the step amplitude in the model update, lets recall the definition of the scalar reflection coefficient and write it in terms of the slowness ($s_0 = 1/c_0$ and $s_1 = 1/c_1$) of each medium as,

$$R = \frac{c_1-c_0}{c_1+c_0} = \frac{1-c_0/c_1}{1+c_0/c_1} = \frac{1-s_1/s_0}{1+s_1/s_0}. \tag{30}$$

Solving for s_1/s_0,

$$\frac{s_1}{s_0} = \frac{(1-R)}{(1+R)}. \tag{31}$$

Expanding the denominator of equation 31 as a Taylor series,

$$\frac{s_1}{s_0} = (1-R)(1-R+R^2-R^3...), \tag{32}$$

and keeping terms up to the first order results

$$\frac{s_1}{s_0} \approx (1-R)(1-R), \tag{33}$$

$$\approx 1-2R+R^2. \tag{34}$$

Then, squaring both sides of equation 34 and collecting again terms up to the first order results

$$\frac{s_1^2}{s_0^2} \approx (1-2R)^2 = 1-4R+R^2, \tag{35}$$

$$\approx 1-4R. \tag{36}$$

Solving for s_1^2 reads

$$s_1^2 \approx s_0^2 - 4s_0^2 R. \tag{37}$$

Finally, the squared-slowness change that must happen to get the reflection coefficient R is given by

$$\Delta S \approx s_1^2 - s_0^2 = -4s_0^2 R, \tag{38}$$

$$\approx -\frac{4R}{c_0^2}. \tag{39}$$

Comparing equation 29 and 39 it can be concluded that the inverse Hessian produces a model update which is equal to the change in slowness associated with the reflection coefficient R, up to a first-order approximation. While the role of the gradient is to provide the location of the interface through a Heaviside function, the "spiky" character of the approximate inverse Hessian is focusing and correcting the amplitude of the gradient, leading to a first order approximation of the actual slowness change.

1D Full-Hessian

For finding an analytical expression for full-Hessian, the $H_1(\mathbf{r}', \mathbf{r})$ term given in equation 8 should be computed. This term includes information about the residual wavefield $\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega)$ into the Hessian computation. The complex conjugate of the residual wavefield for a 1D scalar medium FWI can be expressed as (Innanen, 2013)

$$\delta P^*(0,0,\omega) = -R_1 \frac{e^{-i2k_0 z_1}}{i2k_0}. \tag{40}$$

Rewriting equation 8 for the 1D case results

$$H_1(z',z) = -\int d\omega\omega^4[G(0,z,\omega)G(z,z',\omega)G(z',0,\omega) + G(0,z',\omega)G(z',z,\omega)G(z,0,\omega)]\delta P^*(0,0,\omega). \tag{41}$$

Substituting equations 10, 12, 13 and 40 into 41, it can be written as

$$H_1(z',z) = -\int d\omega\omega^4 \left[\frac{e^{ik_0z}}{i2k_0}\frac{e^{ik_0|z-z'|}}{i2k_0}\frac{e^{ik_0z'}}{i2k_0} + \frac{e^{ik_0z'}}{i2k_0}\frac{e^{ik_0|z'-z|}}{i2k_0}\frac{e^{ik_0z}}{i2k_0}\right]\left(-R_1\frac{e^{-i2k_0z_1}}{i2k_0}\right).$$

Since $|z'-z| = |(-1)(z-z')| = |z-z'|$,

$$H_1(z',z) = \frac{R_1}{2^4}\int d\omega 2\left(\frac{\omega}{k_0}\right)^4\left[e^{ik_0(z+z'-2z_1+|z-z'|)}\right]. \tag{42}$$

Defining the variable $Z = (z+z'-2z_1)$ equation 42 can be written as

$$H_1(z',z) = \frac{R_1}{2^3}\int d\omega\left(\frac{\omega}{k_0}\right)^4\left[e^{ik_0(Z+|z-z'|)}\right]. \tag{43}$$

The integral in equation 43 can be solved in a similar way to equation 15, therefore

$$H_1(z',z) = \frac{R_1c_0^5\pi}{2^3}\delta(Z+|z-z'|). \tag{44}$$

As we can see in equation 44, the $H_1(z',z)$ term is a new delta function, which has an absolute value function within it. This result confirms that the role of the full-Hessian is to scale the gradient function, after integrating it over depth. Hence, the depth of the interface predicted by the gradient function will not be changed by the Hessian.

Lets take one more step forward and consider $z > z'$. Then

$$H_1(z',z) = \frac{R_1c_0^5\pi}{2^3}\delta(Z+z-z'), \tag{45}$$

$$= \frac{R_1c_0^5\pi}{2^3}\delta(z+z'-2z_1+z-z'), \tag{46}$$

$$= \frac{R_1c_0^5\pi}{2^3}\delta(2(z-z_1)), \tag{47}$$

Using the scaling properties of the delta function,

$$H_1(z',z) = \frac{R_1c_0^5\pi}{2^3}\left(\frac{1}{2}\right)\delta(z-z_1). \tag{48}$$

$$= \frac{R_1c_0^5\pi}{2^4}\delta(z-z_1). \tag{49}$$

It can also be shown, following the same steps as before, that for $z' > z$ the $H_1(z, z')$ term reduces to

$$H_1(z', z) = \frac{R_1 c_0^5 \pi}{2^4} \delta(z - z_1). \tag{50}$$

For both cases (equations 49 and 50) the effect of the residual-dependent term of the Hessian is to scale the gradient function just at the depth of the interface. Compared with the approximate-Hessian (equation 18), the $H_1(z', z)$ includes information about R_1 which was not present before.

Substituting equations 18 and 44 into equation 7 the full-Hessian can be written as

$$H(z', z) = \frac{R_1 c_0^5 \pi}{2^3} \delta(Z + |z - z'|) + \frac{c_0^5 \pi}{2^4} \delta(z' - z). \tag{51}$$

Comparing equation 51 and 18 reveals that the full-Hessian is now carrying out additional information just at the depth of the interface (z_1). From equation 49, the scaling to be applied by the $H_1(z', z)$ term is R_1 times the scaling of $H_2(z', z)$. Hence, it can also be concluded that in the presence of large reflection coefficients the residual-dependent term of the full-Hessian will have an important effect on the FWI.

Developing an analytical expression for the inverse of the full-Hessian is one of the goals of an ongoing research project. However, we may expect it to keep the feature mentioned above regarding the localization of the additional scaling. Tarantola (2005) states that full-Hessian may become different to the approximate-Hessian, not only when the residuals are large but when the non-linearities of the forward problem are large. Therefore, it can be expected that using the full-Hessian may have an effect similar to including high order terms into the approximation for computing the magnitude of the model update, in contrast to the first order approximation given by the approximate-Hessian.

Numerical experiments

Hessian in matrix-vector form

Computing the approximate- or the full-Hessian numerically may have a high computational cost, especially for multi-parameter or multi-dimensional cases. The approximate-Hessian requires m forward modellings, while the full-Hessian requires m^2 additional forward modellings, where m is the number of model cells (Pratt et al., 1998).

Mathematical expressions for the numerical computation of the Hessian are better understood in their matrix form. In this setting, the Hessian can be computed as (Pratt et al., 1998)

$$\mathbf{H} = \mathrm{Re}\left\{\mathbf{J}^{\mathrm{t}}\mathbf{J}\right\} + \mathrm{Re}\left\{\left[\left(\frac{\partial}{\partial p_1}\mathbf{J}^{\mathrm{t}}\right)\delta\mathbf{d}^* \quad \left(\frac{\partial}{\partial p_2}\mathbf{J}^{\mathrm{t}}\right)\delta\mathbf{d}^* \quad \ldots \quad \left(\frac{\partial}{\partial p_m}\mathbf{J}^{\mathrm{t}}\right)\delta\mathbf{d}^*\right]\right\}, \tag{52}$$

where $\mathbf{J}$ is the Frechét derivative matrix, the elements of which are given by

$$J_{ij} = \frac{\partial u_i}{\partial p_j}, \qquad i = (1, 2, \ldots, n); \qquad j = (1, 2, \ldots, m). \tag{53}$$

$\delta\mathbf{d}^*$ is the complex conjugate of the residuals vector with elements

$$\delta d_i = u_i - d_i, \qquad i = (1, 2, \ldots, n), \tag{54}$$

where u_i is the model response at receiver location i, d_i is the modelled data at the same location, and p_j is the j-th model parameter.

As in equation 52, the full-Hessian can be split in two terms. The approximate Hessian

$$\mathbf{H_a} = \mathrm{Re}\left\{\mathbf{J}^{\mathrm{t}}\mathbf{J}\right\}, \tag{55}$$

and the residual-dependent term,

$$\mathbf{R} = \mathrm{Re}\left\{\left[\left(\frac{\partial}{\partial p_1}\mathbf{J}^{\mathrm{t}}\right)\delta\mathbf{d}^* \quad \left(\frac{\partial}{\partial p_2}\mathbf{J}^{\mathrm{t}}\right)\delta\mathbf{d}^* \quad \ldots \quad \left(\frac{\partial}{\partial p_m}\mathbf{J}^{\mathrm{t}}\right)\delta\mathbf{d}^*\right]\right\}. \tag{56}$$

Equations 55 and 56 are equivalent to the expressions in equation 9 and 8 respectively. Equation 55 can be interpreted as the zero-lag cross-correlation between the first-order partial derivatives of the wavefields computed at different model cells. Similarly, equation 56 represents the cross-correlation of the second-order partial derivative of the wavefields and the data residuals (Pratt et al., 1998).

Once the Hessian has being computed the model update δp is given by

$$\delta\mathbf{p} = -\mathbf{H}^{-1}\Delta_p E, \tag{57}$$

where $\Delta_p E$ is the gradient of the objective function $E(\mathbf{p})$ and can also be computed in terms of $\mathbf{J}$ and the data residuals as

$$\Delta_p E = \frac{\partial E}{\partial \mathbf{p}} = \mathrm{Re}\{\mathbf{J}^t\delta\mathbf{d}^*\}. \tag{58}$$

Full-Newton and quasi-Newton 1D FWI

In equations 55 and 56, the key term for computing the Hessian is the partial derivative seismograms. Here we use a first-order differencing scheme,

$$\frac{\partial u}{\partial p_j} = \frac{u(p_{0,j} + \Delta p) - u(p_{0,j})}{\Delta p}, \tag{59}$$

where $u(p_{0,j})$ is the forward modelled seismogram using the initial parameter model value p_0 and $u(p_{0,j} + \Delta p)$ is the modelled seismogram after "perturbing" the initial parameter model. This computation has to be done after perturbing each j-th cell of the parameter model. The same first order differencing scheme may be used for computing the derivatives of the partial derivative seismograms, i.e. the second derivative of the wavefields respect to each parameter model.

Figure 2 shows the initial model and the actual model used for the tests. The initial model has a constant velocity chosen to match the velocity of the first layer. Synthetic

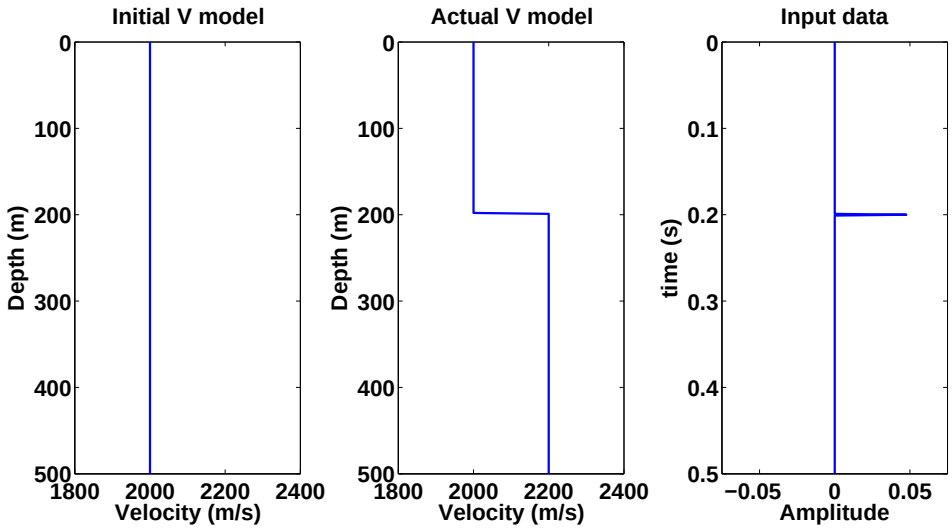

FIG. 2. (left) Initial velocity model used for the inversion (velocity was chosen to be constant with a value equal to the velocity of the first layer ($c_0 = 2,000m/s$)). (middle) "Actual" velocity model used for computing the input data shown at the right.

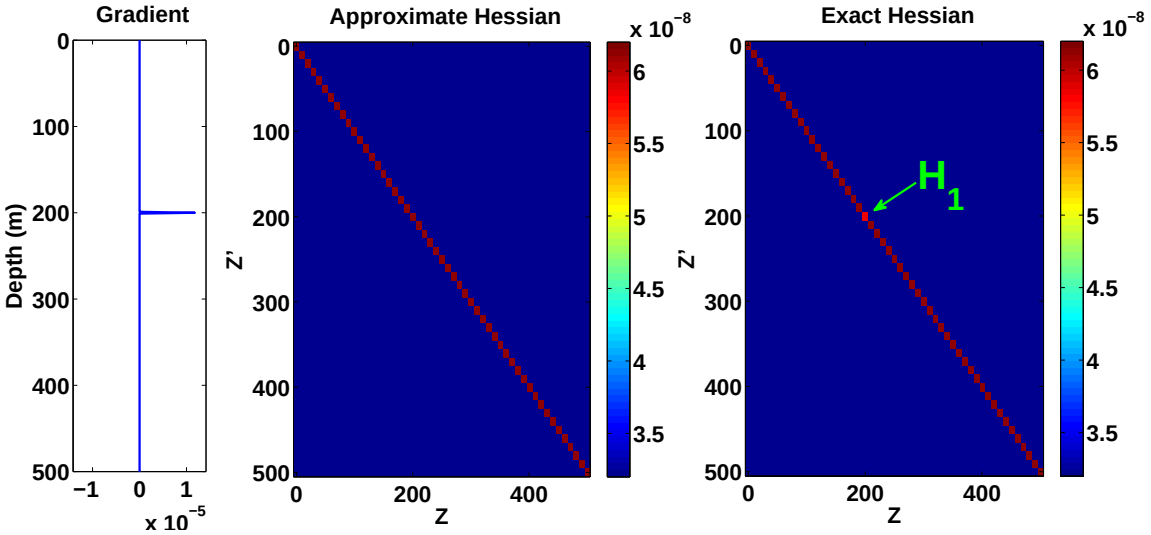

FIG. 3. (left) Gradient of the objective function computed by the means of the cross-correlation of the derivative matrix $\mathbf{J}$ and the data residuals. (middle) Approximate-Hessian. (right) Full-Hessian. The small change in the diagonal value at 200m depth was introduced by the residual dependent term of equation 52. The colour scale has been clipped to highlight the presence of the change in the diagonal values

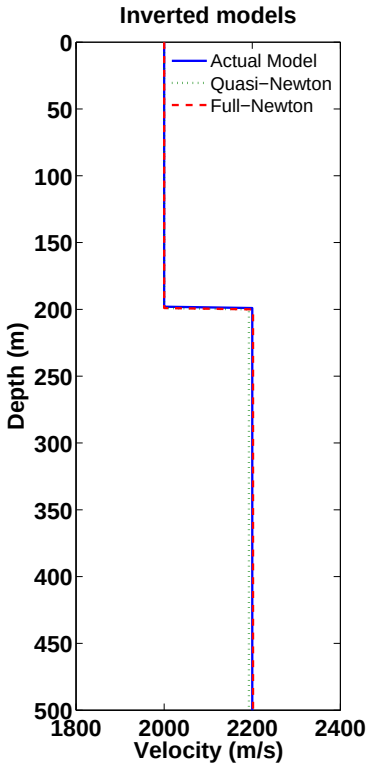

FIG. 4. Inversion results after using both the approximate-Hessian (quasi-Newton method) and the full-Hessian (full-Newton method).

data was computed as full-band seismograms hence the input trace is a spike scaled by the reflection coefficient R_1 and located at time $t_1 = 2z_1/c_0$.

Following the relationships established in equations 55, 56 and 58, we computed the gradient and both the approximate- and full-Hessians (Figure 3). As it was expected, the gradient is a spike located at the depth of the interface. Integration of this spike respect to depth leads to the "step" function that was used in the analytical derivation of the Hessian (equation 25). The approximate-Hessian is a diagonal matrix and the exact Hessian looks very close to it but it has additional information at the depth of 200 m. From the analytical derivation (equations 44 and 51) it is known that the residual dependent term in the full-Hessian provides additional information at the depth of the interface, which is shown in figure 3 as a small change in the diagonal value at z_1 depth.

From equation 3, the model update should be given by the multiplication of the inverse Hessian and the gradient vector. However, computing the inverse of the Hessian is not needed since there are numerical methods for directly computing this product without inverting the matrix.

Figure 4 shows the inverted models using both versions of the Hessian. The inverted velocity values are very close to the actual model ($c_{1,QN} = 2,192m/s; c_{1,FN} = 2,202m/s$). While the full-Newton FWI gave almost a perfect match with a relative error of just 0.09%, the quasi-Newton gave a relative error of 0.35%, in just one step. This shows the power

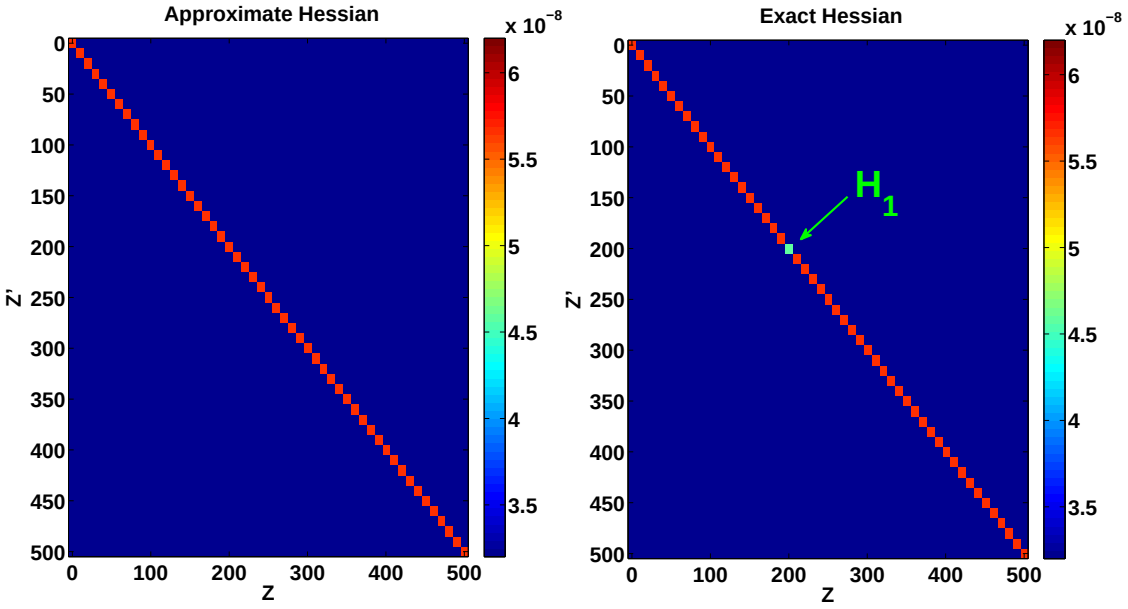

FIG. 5. Approximate-Hessian (left) and full-Hessian (right) after increasing the velocity contrast up to 50%. The change in the diagonal at 200m has now a larger value than in figure 3. Colour bar has been clipped and set at the same range than figure 3

of the Hessian in the search for the minimum of the objective function. Compared to a gradient-based FWI where, after computing the gradient, a line search must be performed to get the proper step-length, the Hessian is directly producing a model update which leads the inversion closer to the minimum of the objective function in just one step.

The second important result obtained from the analytical derivation, was the role of the full-Hessian in the presence large velocity contrasts. Figure 5 shows both Hessians for a model with a contrast of 50% the background velocity (i.e. $c_0 = 2,000m/s; c_1 = 3,000m/s$). Although both Hessians look very similar, the change in the diagonal value located at the interface depth, is larger after increasing the contrast. This effect is related with the dependency on R_1 of the scaling produced by the residual-dependent term in the full-Hessian.

Figure 6 shows the inverted models after considering the 50% increase in velocity mentioned above and an additional result for an increase of 100% in the velocities. The quasi-Newton FWI result shows a consistent underestimation of the velocity of the second layer by a relative error of 5% ($c_{1,QN} = 2,840m/s$), for the 50% velocity contrast and up to 13% ($c_{1,FN} = 3,467m/s$), when the contrast is 100% larger. On the other hand, despite the full-Newton FWI also shows an increase in its relative error, it does so in a much lower magnitude. The relative errors obtained using the full-Hessian were of 1.27% ($c_{1,FN} = 3,038m/s$) for the 50 % velocity contrast, and 2.8% ($c_{1,FN} = 4,112m/s$) for the 100% velocity contrast. Results for both contrast levels are still very good for being just the first step into a larger loop of iterations.

The behaviour of the relative error in the estimation of c_1 for a wide range of velocity contrasts can be seen in Figure 7. The error obtained using the full-Newton inversion is consistently lower than the error of the quasi-Newton method. While the quasi-Newton

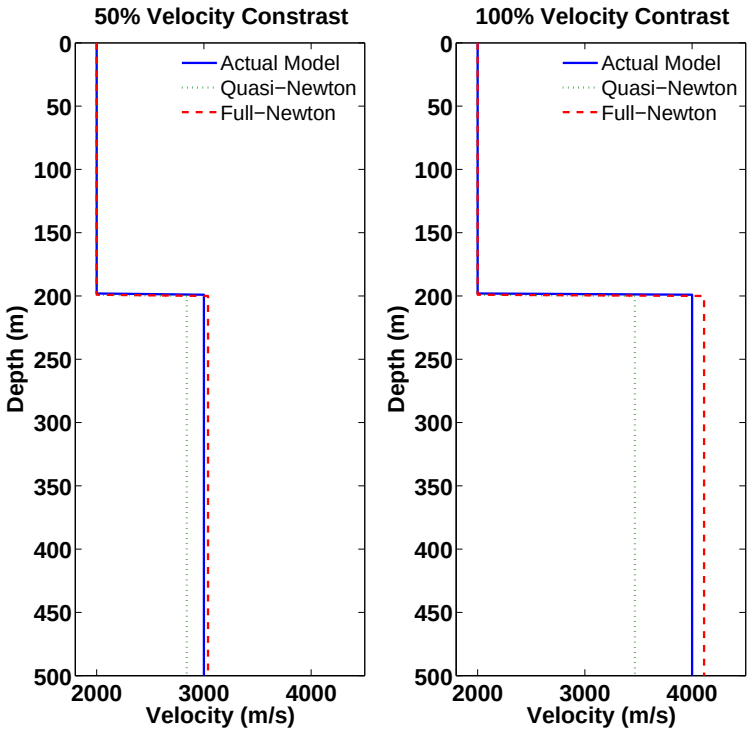

FIG. 6. Inversion results after increasing the velocity the velocity contrast up to 50% (left) and 100% (right).

method tends to increasingly underestimate the value of c_1 the full-Newton FWI is very close to the actual c_1 value, slightly overestimating its value for velocity ratios up to 4.8, before starting to underestimate it. This result may have an important effect in the convergence of the FWI to the actual c_1 value. Since the velocity updates given by the quasi-Newton method are smaller than the updates given by the full-Newton inversion it will be needed a larger number of iterations to achieve the actual velocity value. Despite the high computational cost of computing the full-Hessian, in settings with large velocity contrasts it will lead to a more accurate velocity estimation. Moreover, the ability of the full-Hessian for taking into account the non-linear part of the amplitude changes makes it a very good candidate for inverting seismic data under the presence of large velocity contrasts or when AVO information is important.

CONCLUSIONS

The analytical development made in this work showed that the approximate-Hessian provides the proper scaling for obtaining a very close match to the actual velocity change in just one step. Comparing the model update given by a quasi-Newton step with a series expansion of the reflection coefficient, it can be concluded that the update given by the approximate-Hessian is exact up to a first-order approximation of the velocity change related with the actual reflection coefficient. On the other hand, by the means of including information about the residuals, the full-Hessian is able to provide additional scaling at the actual depth of the interface. The numerical experiments confirmed the expected effect of the full-Hessian regarding its role when the non-linearity of the forward problem

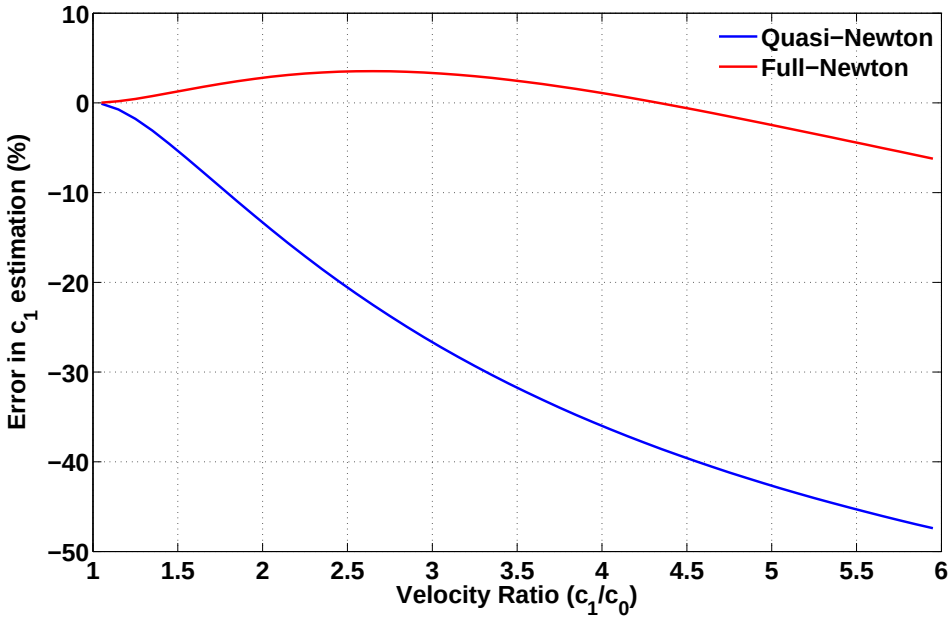

FIG. 7. Error in the estimation of c_1 for the quasi- and full-newton FWI, based on the relative magnitude of the velocity contrasts.

becomes important. On the presence of a large velocity contrast, the full-Newton FWI was able to give a very close match to the actual velocity model in just one step. Despite the approximate-Hessian is able to give a very good match when velocity contrasts are small, it starts to fail when velocity changes are very large. Moving to a 2D scenario, is needed to understand the role of the Hessian when large AVO effects are present in the data.

There is still much work to do in terms of the time-efficient computation of the Hessian for large datasets. The requirement of a huge number of forward modellings makes it very costly in a computational sense, particularly when multidimensional or multiparemeter FWI comes into the scene. However, computational power has shown to increase exponentially with time, making the use of the full-Hessian a very good option for FWI in a short to medium term.

ACKNOWLEDGEMENTS

The authors would like to acknowledge to Faranak Mahmoudian for helping in the edition of this paper and to CREWES sponsors for providing the financial support for this research.

REFERENCES

Innanen, K. H., 2013, GOPH 673: Inverse theory and application II: Course Notes. University of Calgary.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Proceedings of the international conference on inverse scattering, theory and applications: Society of Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

Margrave, G., Yedlin, M., and Innanen, K., 2011, Full waveform inversion and the inverse hessian: CREWES

Research Report, **23**, 77.1–77.13.

Pratt, G., Shin, C., and Hicks, 1998, Gauss-newton and full newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, No. 2, 341–362.

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

Tarantola, A., 2005, Inverse Problem Theory and Methods for Model Parameter Estimation: Society for Industrial and Applied Mathematics.

Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, No. 6, WCC1–WCC26.