Enhancing Full Waveform Inversion with a deep-learning approximated inverse Hessian: A field data application

Mustafa Alfarhan, Matteo Ravasi, Fuqiang Chen, and Tariq Alkhalifah*
King Abdullah University of Science and Technology (KAUST), Saudi Arabia

SUMMARY

In image-domain Least-Squares Migration (LSM), a demigration-migration approach can be employed to identify non-stationary local filters that approximate the inverse Hessian operator. Such filters can be subsequently applied to the original migrated image to compensate for uneven illumination, undo the effect of geometrical spreading, and enhance its resolution. A similar approach can be applied at each iteration of Full Waveform Inversion (FWI) by taking the FWI gradient as input to the demigration-migration process. By doing so, a more balanced update can be constructed whereby both the shallow and deep parts of the model are updated with similar strength. Following our recent line of work, we propose here to approximate the effect of the inverse Hessian with a neural network, which is trained to map the doubly migrated gradient into the FWI gradient. The trained network is later applied to the gradient itself to produce an improved FWI model update. Compared to conventionally used non-stationary local filters, the network can be trained only once (at the first FWI iteration) and cheaply fine-tuned at any subsequent iteration. In this work, we apply the proposed methodology to a challenging field dataset, namely the 2010 ocean-bottom cable Volve dataset. As commonly done in FWI, our approach is naturally embedded into a multi-scale approach where three different frequency bands are subsequently inverted and the network used to approximate the inverse Hessian is re-trained at each outer iteration. When compared to state-of-the-art quasi-Newton methods, the model obtained using our approach is shown to produce images of superior quality and flatter as well as more focused angle gathers.

INTRODUCTION

Full Waveform Inversion (FWI), (Lailly and Bednar, 1983; Tarantola, 1984), has gained significant traction over the past two decades due to its ability to construct high-resolution subsurface models by leveraging the information contained in the full seismic waveform. In comparison, traditional methods for velocity model building mainly utilize travel times and discard valuable information contained in the amplitude and phase of seismic waves. FWI relies on a forward modeling engine to obtain simulated data from an initial model, which is directly compared to the observed data. Such a comparison is carried out using a misfit function such as the ℓ^2 norm to calculate an update for the model. The update includes the first and second derivatives of the misfit function with respect to the model parameters, also known as the gradient and the Hessian, respectively. The inverse of the Hessian scales the gradient to account for the losses due to geometrical spreading as well as transmission. Thus, approximating the inverse of the Hessian can speed up the convergence of FWI and result in an enhanced inverted model.

Depending on how much computation we can afford, we content with an approximation of the update ranging from only using the gradient to including the full Hessian where the latter is the bottleneck in the computation. One way to include the inverse Hessian in the FWI update is to approximate it using second-order methods such as the Gauss-Newton (GS) method (Pratt et al., 1998), quasi-Newton methods such as the limited-memory BFGS algorithm (Liu and Nocedal, 1989), or the truncated-Newton methods (AlTheyab et al., 2013; Métivier et al., 2014). The inverse of the Hessian, also, plays an important role in enhancing seismic images where it can be viewed as a deblurring operator which can improve the resolution of migrated images. Due to the linearity of the migration operator, image processing techniques can be utilized to approximate the inverse Hessian such as non-stationary matching filters as done in (Guitton, 2004), deblurring filters (Aoki and Schuster, 2009), and point spread functions (PSFs) (Valenciano, 2008). Recently, deep learning approaches have been used to approximate the action of the inverse Hessian in seismic imaging and have been shown to be faster than state-of-the-art methods (Liu et al., 2022; Kumar et al., 2022; Vasconcelos et al., 2022).

In our recent work, we showed that the idea of using deep learning to approximate the inverse Hessian can also be used in FWI (Alfarhan et al., 2023). We demonstrated the network's capability of speeding up the convergence of FWI with synthetic examples. In this work, we assess the validity of our approach on the Volve field dataset. The Theory section reviews our proposed method; this is followed by a Numerical Examples section providing a description of the Volve dataset and the sequence of pre-processing steps performed prior to FWI, and presenting the results of our method on this dataset. Lastly, the Discussion section highlights the advantages and limitations of our method, and we end with conclusion remarks in the Conclusion.

THEORY

The FWI model update, as outlined in Equation 1, involves calculating both the gradient (first derivative) and the Hessian (second derivative) with respect to the misfit function J relative to the current model $\mathbf{m}_o$.

$$\delta \mathbf{m} = -\left(\frac{\partial^2 J(\mathbf{m}_o)}{\partial \mathbf{m}^2}\right)^{-1} \frac{\partial J(\mathbf{m}_o)}{\partial \mathbf{m}} \tag{1}$$

Note that whilst any differentiable function can be used in theory, in this work we select the ℓ^2 norm. To approximate the

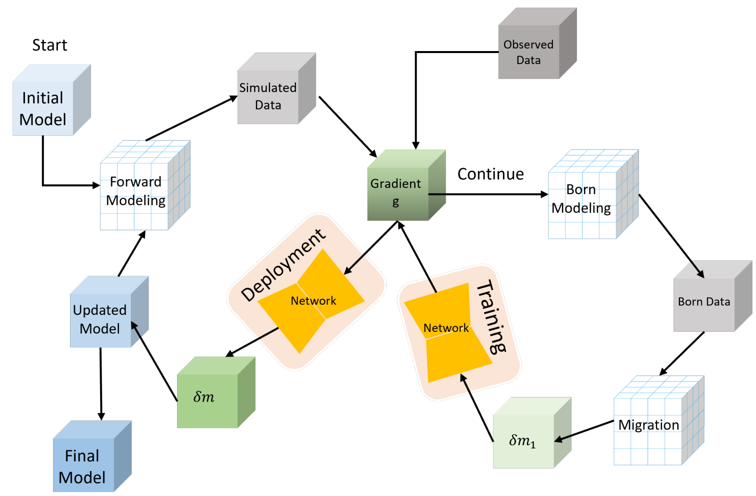

Figure 1: A depiction of the proposed method.

Hessian, we rely on linearizing the wave equation, which allows us to view the FWI gradient $\mathbf{g}$ as the migration of the data residual $\Delta\mathbf{d}$ (the difference between the modeled and observed data) using the adjoint of Born modeling $\mathbf{L}^T$:

$$\mathbf{g} = \mathbf{L}^T \Delta\mathbf{d} \tag{2}$$

Furthermore, the residual data $\Delta\mathbf{d}$ can be expressed as the demigrated Earth perturbation $\delta\mathbf{m}$ through the Born modeling operator $\mathbf{L}$ as follows:

$$\Delta\mathbf{d} = \mathbf{L}\delta\mathbf{m} \tag{3}$$

Therefore, the combination of equations 2 and 3, as shown in equation 4, reveals that the FWI gradient, $\mathbf{g}$, is connected to the Earth perturbation $\delta\mathbf{m}$ through $(\mathbf{L}^T\mathbf{L})$ or the Hessian.

$$\mathbf{g} = \mathbf{L}^T\mathbf{L}\delta\mathbf{m} \tag{4}$$

By applying the Hessian $(\mathbf{L}^T\mathbf{L})$ to the FWI gradient, we can obtain another migrated image:

$$\delta\mathbf{m}_1 = \mathbf{L}^T\mathbf{L}\mathbf{g} \tag{5}$$

Hence, we can consider the FWI gradient, $\mathbf{g}$, and the migrated image, $\delta\mathbf{m}_1$, to be linked through the inverse Hessian $(\mathbf{L}^T\mathbf{L})^{-1}$. Consequently, using a neural network to learn the mapping from $\delta\mathbf{m}_1$ to $\mathbf{g}$ is approximately equivalent to the inverse Hessian that operates on $\mathbf{g}$ to get $\delta\mathbf{m}$, the full model update in FWI. As a result, FWI will converge faster, and we can repeat the same process while leveraging transfer learning to reduce the training burden of the neural network. Figure 1 illustrates the proposed method visually.

NUMERICAL EXAMPLES

The Volve oil field, located in the North Sea off the coast of Norway, was decommissioned in 2016. In 2018, Equinor publicly released the well, production, and seismic data for research and educational purposes. In this work, we aim to apply the proposed methodology to a 2d line of the ocean-bottom cable (OBC) dataset composed of 110 sources and 180 receivers sampled every 50 and 25 m, respectively. Given one of the key assumptions of our methodology (i.e., Born approximation), the dataset is initially pre-processed to remove any free-surface effect. More specifically, the pressure and vertical particle velocity components are used to obtain the up- and down-going separated wavefields, which are subsequently used as input to a step of multi-dimensional deconvolution. As a result the processed data presents sources and receivers co-located at the seafloor. We refer to (Ravasi et al., 2015, 2016, 2022) for a more comprehensive description of these processing steps. Figure 2 presents a 2D slice of the tomographic velocity model available as part of the open Volve dataset, highlighting the line of sources and receivers chosen for this research in red, and its smoothed counterpart employed as the initial model for FWI.

We compare our approximate method using an approximate inverse Hessian with two commonly used methods to approximate the inverse Hessian: the Barzilai-Borwein (BB) (Barzilai and Borwein, 1988) solver, which is a gradient-based method and L-BFGS (Liu and Nocedal, 1989) which is a quasi-Newton method. Given that the inner working of these algorithms is different (and therefore the number of wave equation solves used at each iteration may differ) , we decide to use the number of wave equation solves as ametric to compare the convergence speed of these three methods. Additionally, in all cases FWI is performed in a multi-scale fashion using a maximum frequency of 4, 7, and 10 Hz for the three outer iterations. As shown in Figure 3, L-BFGS lags in speed compared to the other two approaches and reaches an overall higher data residual. On the other hand, our method converges slightly faster than the BB method at both the 4 and 10 Hz scales. Figure 4 shows the final updated model for the three approaches. The velocity is overestimated in the shallow and deep parts by the L-BFGS method, whilst the velocity model from the BB method has a stronger update in the shallow part compared to our method final velocity model.

Imaging is further carried out using these three final velocity models to assess their accuracy against the initial model, which is a smoothed version of the already highly accurate tomo-

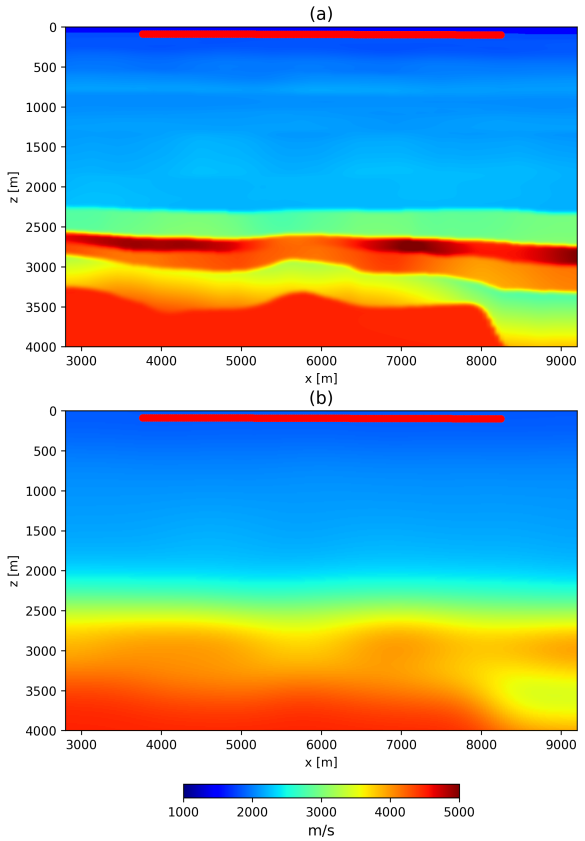

Figure 2: (a) The tomographic velocity model for Volve data and (b) its smoothed counterpart. The red line delineates the arrays of sources and receivers.

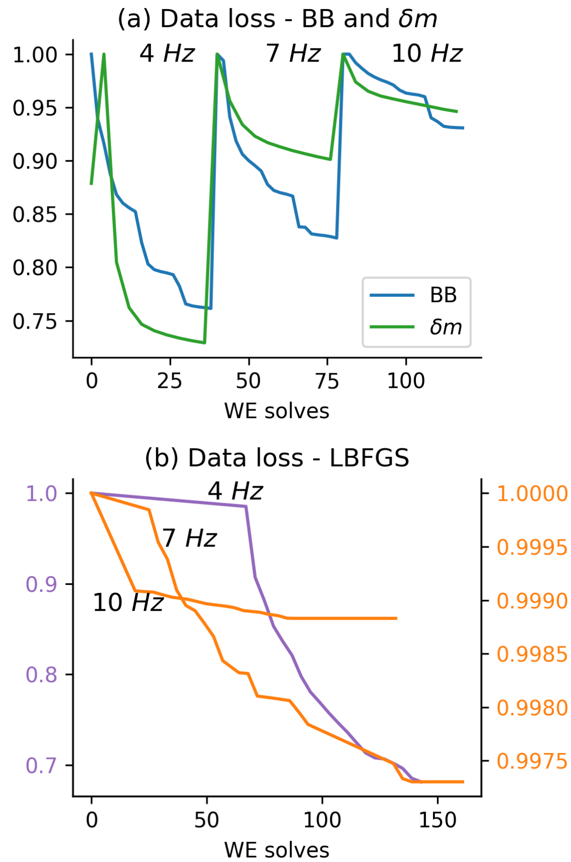

Figure 3: (a) Data loss of running FWI for the BB method and our method δm while (b) represents L-BFGS.

graphic model shown in Figure 2. The Reverse Time Migration (RTM) images using the four velocity models (Figure 5), show that the inverted model with the BB method produces an artifact as indicated by the red arrow in the second panel while the inverted model with L-BFGS (panel (c)) is of overall lower quality when compared to the image from the initial model (panel (a)). In contrast, our method produces a similar image (panel (d)) to the one obtained with the initial model, however better continuity and more focused reflectors are observed in the deeper part of the model as indicated by the red arrows. Lastly, Angle Domain Common Image Gathers (ADCIG) computed at $x = 6.0\ km$ with the same four velocity models used to compute the RTM images are shown in Figure 6. The gather corresponding to our method show similar flatness as the ones corresponding to the initial model with slight improvements (e.g., the one indicated by the red arrow). On the other hand, the gathers related to the BB method and L-BFGS curve down as indicated by the red arrows. Therefore, we can conclude that the inverted model with our approach (and hence the approximated inverse Hessian) is of better quality than the other two approaches.

DISCUSSION

This study introduces an innovative approach to FWI which incorporates an approximation of the inverse Hessian into a neural network designed to transform a Hessian-adjusted gradient into the (FWI) gradient. By employing Born modeling and its adjoint, we integrate the effects of the Gauss-Newton Hessian into the FWI process, with the neural network approximating the inverse Hessian to enhance the update vector at each FWI iteration. We exploit neural networks' adaptability through transfer learning, given the gradual iteration-to-iteration changes in the Hessian's inverse, contrasting with traditional methods that require repetitive recalculations.

However, our approach has limitations, including an approximation to the FWI Hessian that may not accelerate convergence as effectively as the true inverse, and a training process that does not guarantee extension of the frequency content beyond the training data. Additionally, this method introduces computational demands due to extra modeling-migration steps and network training or fine-tuning at each FWI iteration.

Application to the Volve field dataset demonstrated faster or comparable convergence rates to the BB method and improved model quality. The method is most effective with low frequency data and can be adapted for multi-scale FWI by retrain-

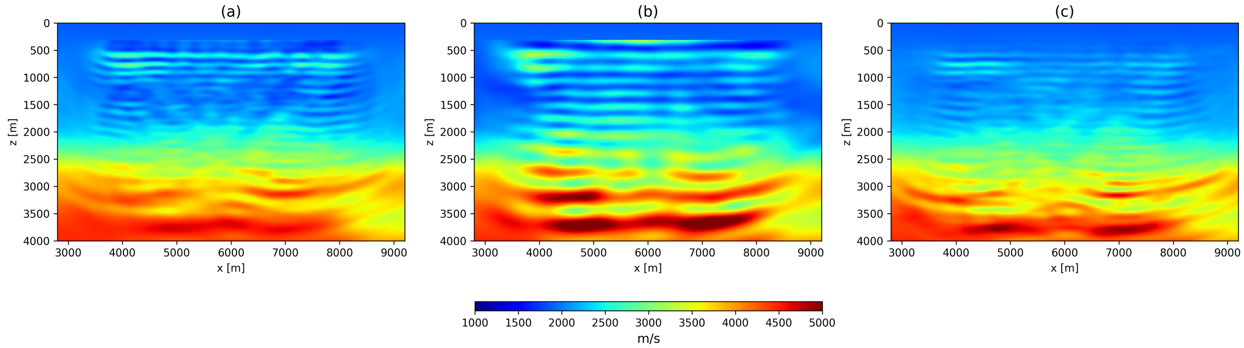

Figure 4: Inverted models: (a) using the BB method, (b) using L-BFGS, and (c) using our method.

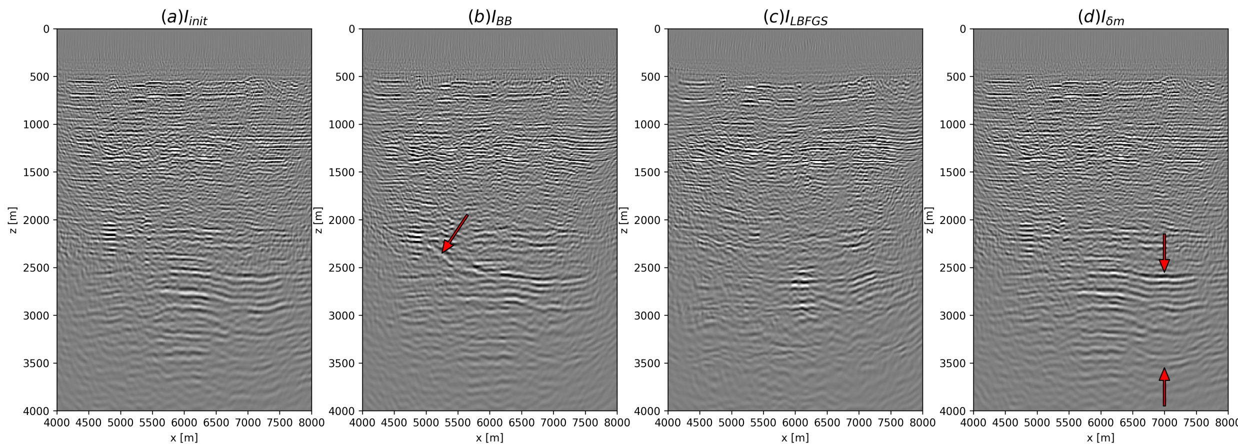

Figure 5: RTM images obtained with: (a) the initial model, (b) the inverted model using the BB method, (c) the inverted model using L-BFGS, and (d) the inverted model using our method.

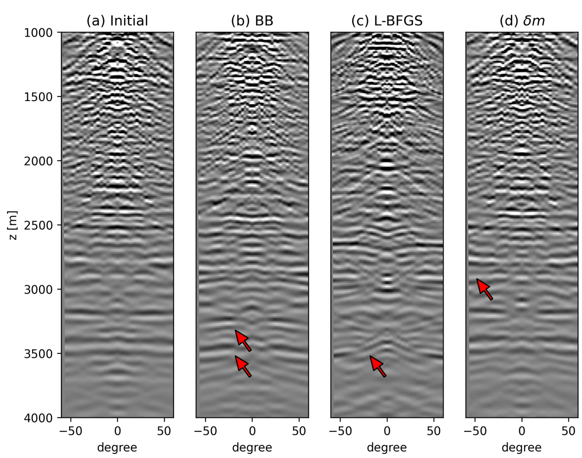

Figure 6: ADCIG obtained at $x = 6.0\ km$ with: (a) the initial model, (b) the inverted model using the BB method, (c) the inverted model using L-BFGS, and (d) the inverted model using our method.

ing the network for each frequency band. Despite the initial computational cost, this inverse Hessian approximation can guide early FWI iterations towards better optimization, offering a practical and cost-effective strategy for enhancing FWI performance.

CONCLUSION

We introduced a deep learning strategy for estimating the inverse Hessian within FWI. This technique involves generating a remigrated image connected to the gradient via the inverse Hessian, utilizing Born modeling and its adjoint. A neural network is subsequently trained to identify the transformation from the remigrated image to the gradient, which, when applied to the gradient, increases its resolution and corrects for illumination discrepancies. Our numerical analyses on the Volve field data demonstrate that gradients refined in this manner contribute to an FWI process characterized by more robust convergence and enhanced quality of the inverted model.

ACKNOWLEDGMENTS

The author thanks KAUST and the DeepWave Consortium sponsors for supporting this research, as well as Equinor and the Volve license partners for releasing the field dataset.