Using the Hessian of a radial basis formulation for level set inversion

Taylor Dahlke, Biondo Biondi, and Robert Clapp

ABSTRACT

Salt bodies provide complex imaging challenges because of their geometry and reflective properties due to the (often) sharp contrast of wave speed between salt and sediments. Level sets are a useful tool to define and refine discrete boundaries of salt using an implicit surface to describe them. Furthermore, we can represent the implicit surface using a sparse representation based on radial basis functions (RBFs). Using linear operators to map from RBF parameter space to wave speed space, we develop a new formulation of the Full Waveform Inversion (FWI) objective function, and then take the second derivative to get a formulation of its Hessian. We can then solve the corresponding Newton system to find a search direction. The sparse representation offered by the RBF scheme means that a truncated iterative inversion is intrinsically faster due to the large reduction in model parameters that we need to solve for. We demonstrate the efficacy of using the Gauss-Newton approximation of this Hessian, as well as explore the limitations of using the full Hessian formulation for finding a search direction.

INTRODUCTION

Previous work done by Kadu et al. (2016) demonstrated how radial basis functions can be used to help define salt features using level sets. Further, there is a growing field of literature describing the use of level sets as a means to track boundaries for shape optimization schemes (Li et al. (2010), Lewis et al. (2012), Guo and Hoop (2013), Santosa (1996), Osher and Sethian (1988), Burger (2003)), including in the domain of seismic imaging. Our recent work explored the use of the Hessian of a full-grid level set formulation (see Dahlke et al. (2017a)), showing promise as a means to find search directions for a FWI-type work flow. This approach included the inversion of a Newton system that uses a full Hessian formulation of the level set-FWI objective function. For large problems like we find in typical 3D seismic imaging, the large number of model parameters makes inverting this system prohibitively expensive. We introduce a sparse parameterization of the level set problem, which allows us to invert a Newton system that is based on roughly two-orders of magnitude fewer model parameters than before. This speeds up the convergence of iterative methods like conjugate gradient. In this work, we derive this formulation, and then demonstrate on 2D synthetic models using the Gauss-Newton Hessian approximation. After this, we

demonstrate using the full Hessian and discuss the practical limitations and potential benefits of that approach.

DERIVATION

Shape optimization

The first step of this derivation is to describe the model space that we are working with. We will call our velocity model m, which we define as:

$$m(\phi_{i,j}, b_{i,j}) = H(\phi_{i,j})(c_{\text{salt}} - b_{i,j}) + b_{i,j}, \tag{1}$$

where $H(\circ)$ is the Heaviside function, $m(\phi_{i,j}, b_{i,j})$ is the velocity value, $\phi_{i,j}$ is the implicit surface value, and $b_{i,j}$ is the background velocity value at 2D spatial position (i, j). We generalize these parameters for the entire spatial domain (ignoring i, j), and expand this definition with a Taylor series as:

$$m_1 = m_0 + \frac{\partial m}{\partial \phi}\Big|_{m_0} \triangle\phi + \frac{\partial m}{\partial b}\Big|_{m_0} \triangle b + ... \tag{2}$$

This approximation is only valid when the Taylor Series we describe converges with the addition of increasingly higher order terms. For the Heaviside function, this is not the case, since the function is not differentiable in original form. For this reason, we must use a smoothed approximation of the Heaviside function, such as:

$$H(\phi) \approx \frac{1}{2}\left[1 + \frac{2}{\pi}\arctan(\frac{\pi\phi}{\epsilon})\right].$$

By truncating the series in equation 2 and ignoring higher order terms, we can create a linear approximation to this smooth approximation for the perturbation of the velocity model m with respect to ϕ and b:

$$\triangle m \approx \frac{\partial m(\phi_o, b_o)}{\partial \phi}\triangle\phi + \frac{\partial m(\phi_o, b_o)}{\partial b}\triangle b. \tag{3}$$

This can be written as a matrix operation:

$$\triangle m \approx \begin{bmatrix} \frac{\partial m(\phi_o, b_o)}{\partial \phi} & \frac{\partial m(\phi_o, b_o)}{\partial b} \end{bmatrix} \begin{bmatrix} \triangle\phi \\ \triangle b \end{bmatrix}$$
$$\triangle m \approx \begin{bmatrix} \frac{\partial m(\phi_o, b_o)}{\partial \phi} & \frac{\partial m(\phi_o, b_o)}{\partial b} \end{bmatrix} \triangle p,$$

where we define operator D as:

$$\begin{aligned} D &= \begin{bmatrix} \frac{\partial m(\phi_o,b_o)}{\partial \phi} & \frac{\partial m(\phi_o,b_o)}{\partial b} \end{bmatrix} \\ &= \begin{bmatrix} \delta(\phi_o)(c_s - b) & 1 - H(\phi_o) \end{bmatrix}. \end{aligned} \tag{4}$$

This operator D ultimately scales and masks the parameter fields $\triangle\phi$ and $\triangle b$. With this new approximation of the perturbation in our velocity model, the application of our Born operator to our new model parameter space $\triangle p = [\triangle\phi \quad \triangle b]^T$ is given by:

$$\triangle d \approx BD\triangle p.$$

Alternatively, we can find the update gradient for our model parameters by applying the adjoint operation:

$$\triangle p \approx D^T B^T \triangle d.$$

Similarly we can find the application of the Hessian to the search direction as:

$$D^T H D \triangle p \approx -D^T B^T \triangle d. \tag{5}$$

In equation 5, we can substitute H with either the full or Gauss-Newton Hessian. Previous work by Fichtner (2010) shows that the full Hessian of the FWI objective function can be constructed by summing a WEMVA component with the Gauss-Newton component of the Hessian. It is this formulation of the full Hessian application that we use. The method we propose solves equation 5 for $\triangle p$ using a conjugate gradient algorithm.

Sparsifying with Radial Basis Functions

The thesis of the work done in Kadu et al. (2016) is to replace a regular grid parametrization of the implicit surface ϕ with a surface described as an aggregate of many RBFs, resulting in a much sparser model. We build upon this idea by clustering the spatial locations of the radial basis functions around the areas we expect to see updating occur. This allows us to use far fewer RBF parameters to attain a higher resolution around the salt boundary than we would if we used the regular gridding described in Kadu et al. (2016):

$$\phi(\lambda;\epsilon,r) = \sum_{i}^{\mathrm{N}\lambda} \lambda_i \exp^{-(\epsilon r)^2} \tag{6}$$

where λ is the new model parameter, r is the radial distance from the RBF center i, and ϵ controls the sharpness of the RBF taper (constant). Further, we limit the

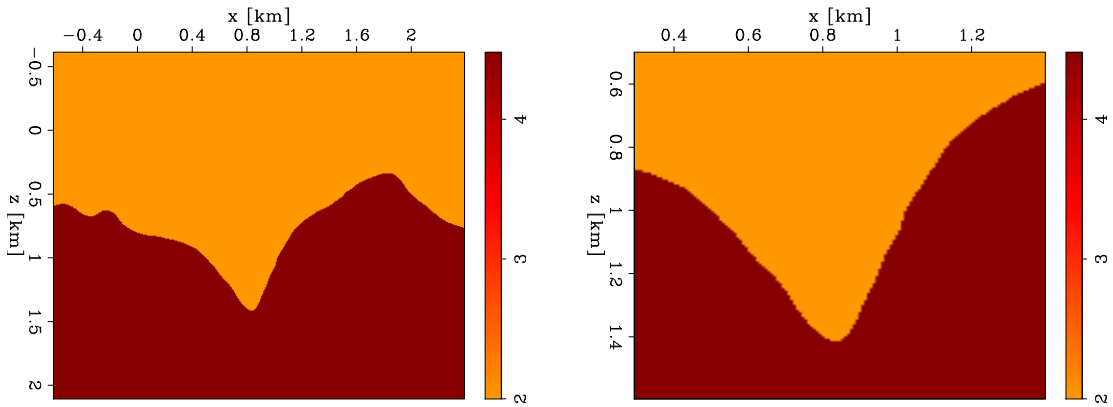
support of each RBF to improve the efficiency of computing the implicit surface ϕ built from the aggregated RBFs. These details and more are described in Dahlke et al. (2017b). In regards to this work however, our operator D must be modified to account for this additional linear transformation:

$$\begin{aligned} D &= \left[\frac{\partial m(\phi_o,b_o)}{\partial \phi}\frac{\partial \phi}{\partial \lambda} \quad \frac{\partial m(\phi_o,b_o)}{\partial b}\right] \\ &= \left[\delta(\phi_o)(c_s - b)\exp^{-(\epsilon r)^2} \qquad 1 - H(\phi_o).\right] \end{aligned} \tag{7}$$

Further, our model space has also changed to be:

$$\triangle p = \begin{bmatrix} \triangle \lambda \\ \triangle b \end{bmatrix}.$$

APPLICATION TO THE SIGSBEE CANYON MODEL



(a) (b)

Figure 1: The full model (a) used for propagation (acquisition geometry from (0,0) to (0,1.75). A close up of the canyon area (b). [**ER**]

Single canyon perturbation

For the first application example, we select a portion of the upper Sigsbee canyon model (Figure 1b). We perturb the left hand side of the canyon (Figure 2) so that we can get secondary scattering against the opposite canyon wall (shown in Figure 3). It is this secondary scattering that the full Hessian is expected to be able to recover, as opposed to the Gauss-Newton Hessian approximation, which relies on only first-order Born scattering. We use this model because it should be able to offer a comparison between the two Hessian formulations of reasonable significance. For this model we use an acquisition geometry of 38 shots evenly spaced, with 230 receivers. We used a

Ricker wavelet with a central frequency of 15 Hz. For this case as well as the double perturbation case, we assume that $\triangle b = 0$, and so invert for a model defined as $\triangle p = \triangle \lambda$.

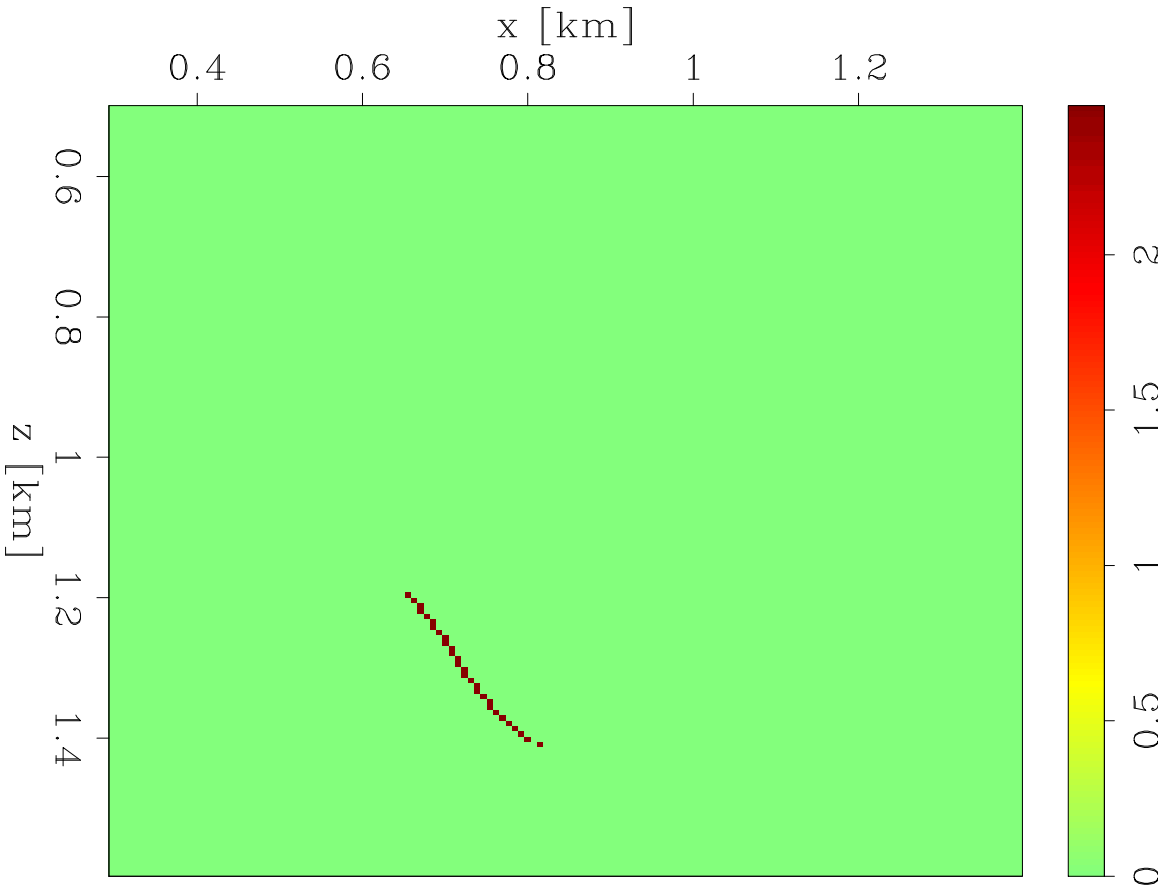


Figure 2: The single canyon perturbation of the Sigsbee model. [**ER**]

Double canyon perturbation

For the second application example, we use the same true model based on the upper Sigsbee salt (Figure 1b). In this case, we perturb both the left and right hand sides of the canyon (Figure 7). This will offer a further complexity to the secondary scattering of the model (shown in Figure 8). The same acquisition geometry and wavelet were used for this example as the first model.

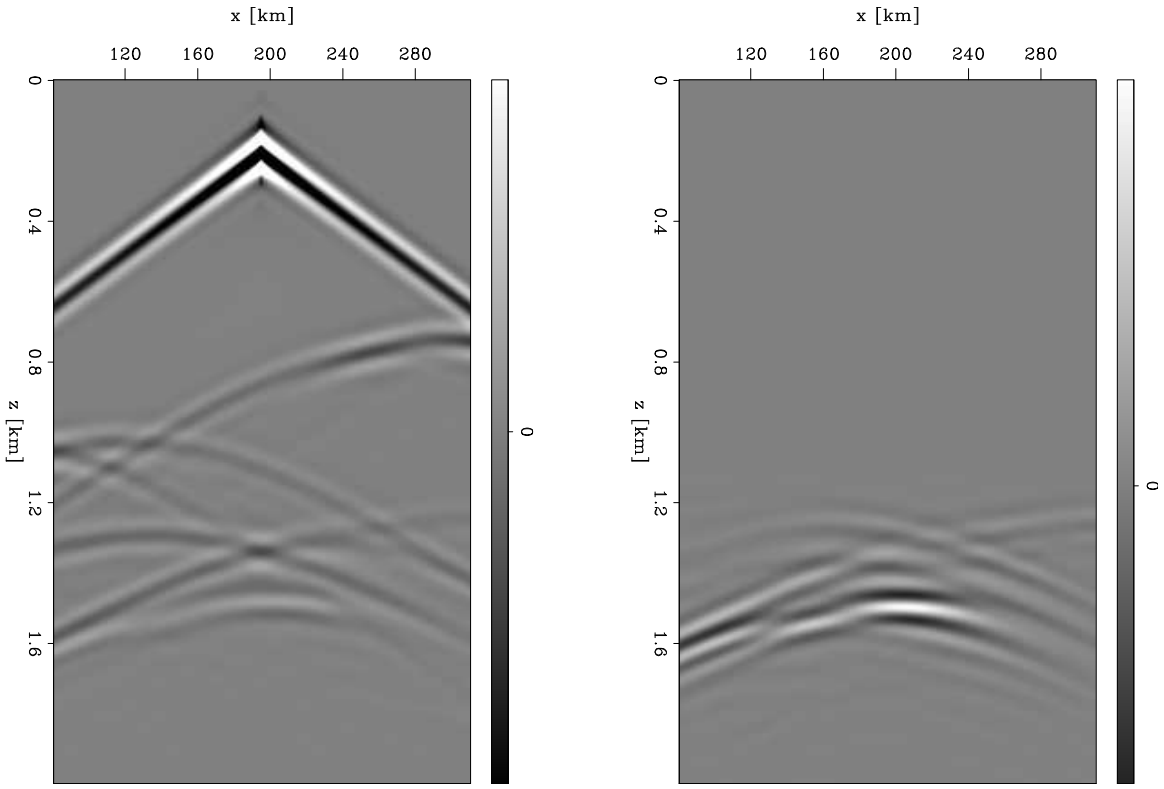


Figure 3: The data generated from the center shot on the true model (left). The center shot residual between the true data and the data generated from an initial guess that had a single canyon side perturbation (right). [**CR**]

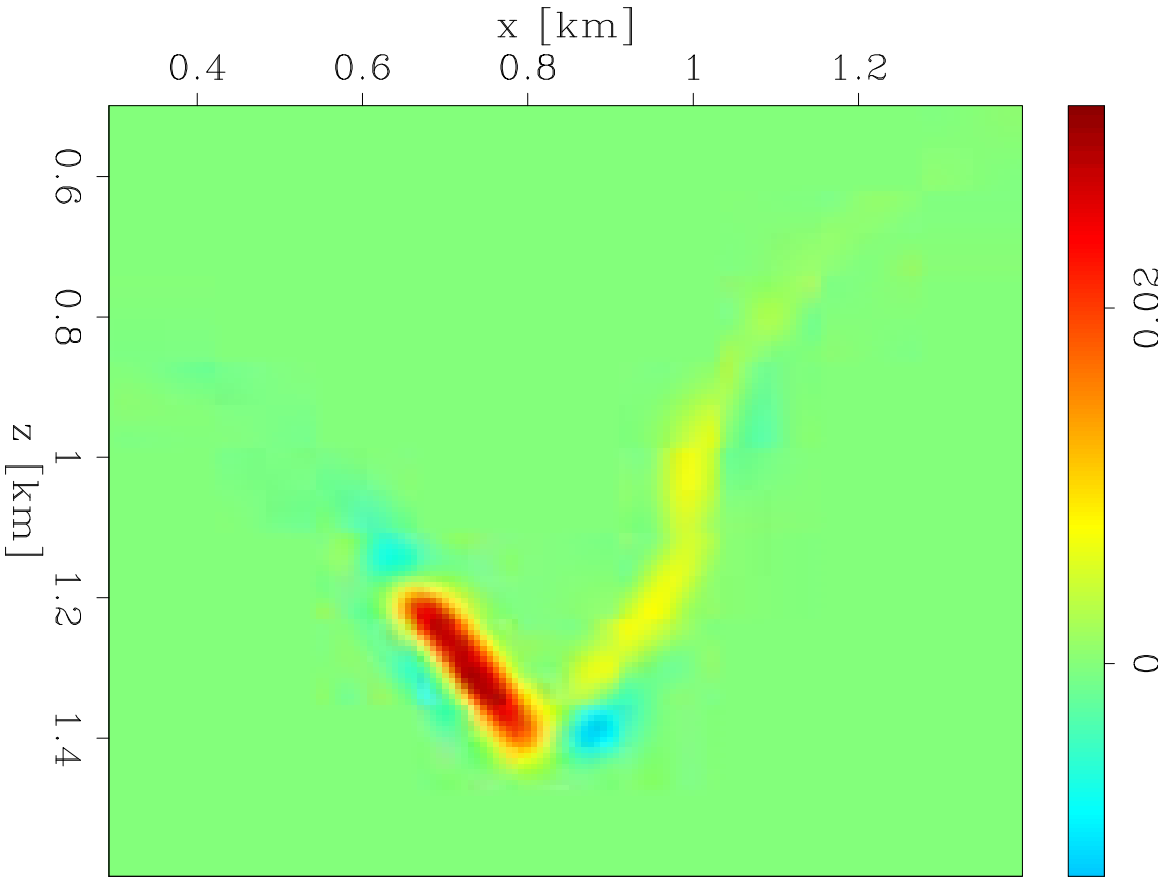


Figure 4: The inversion result after using the Gauss-Newton approximation of the Hessian on the single perturbation model. [**CR**]

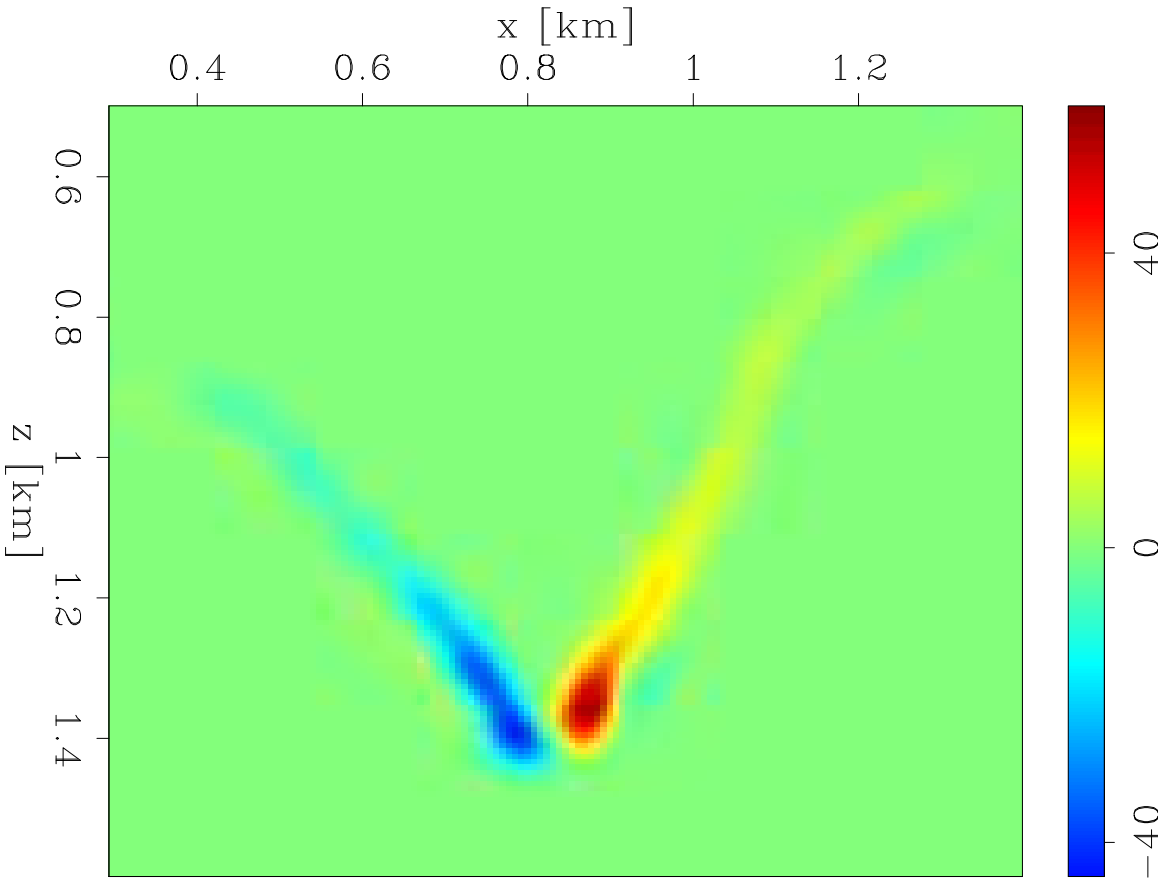


Figure 5: The inversion result after using the full Hessian on the single perturbation model. [**CR**]

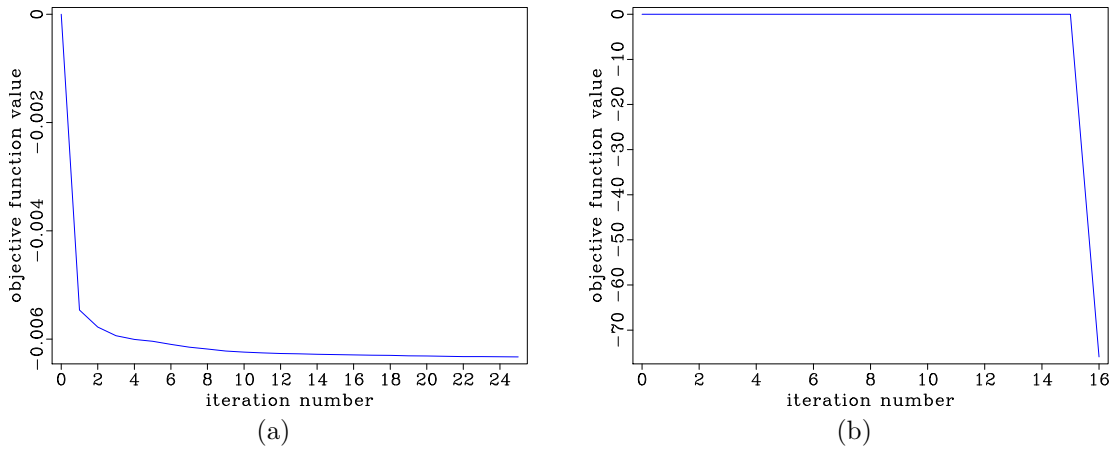


Figure 6: The objective functions from the inversions of the single canyon perturbation model using the Gauss-Newton Hessian (a), and the full Hessian (b). [**CR**]

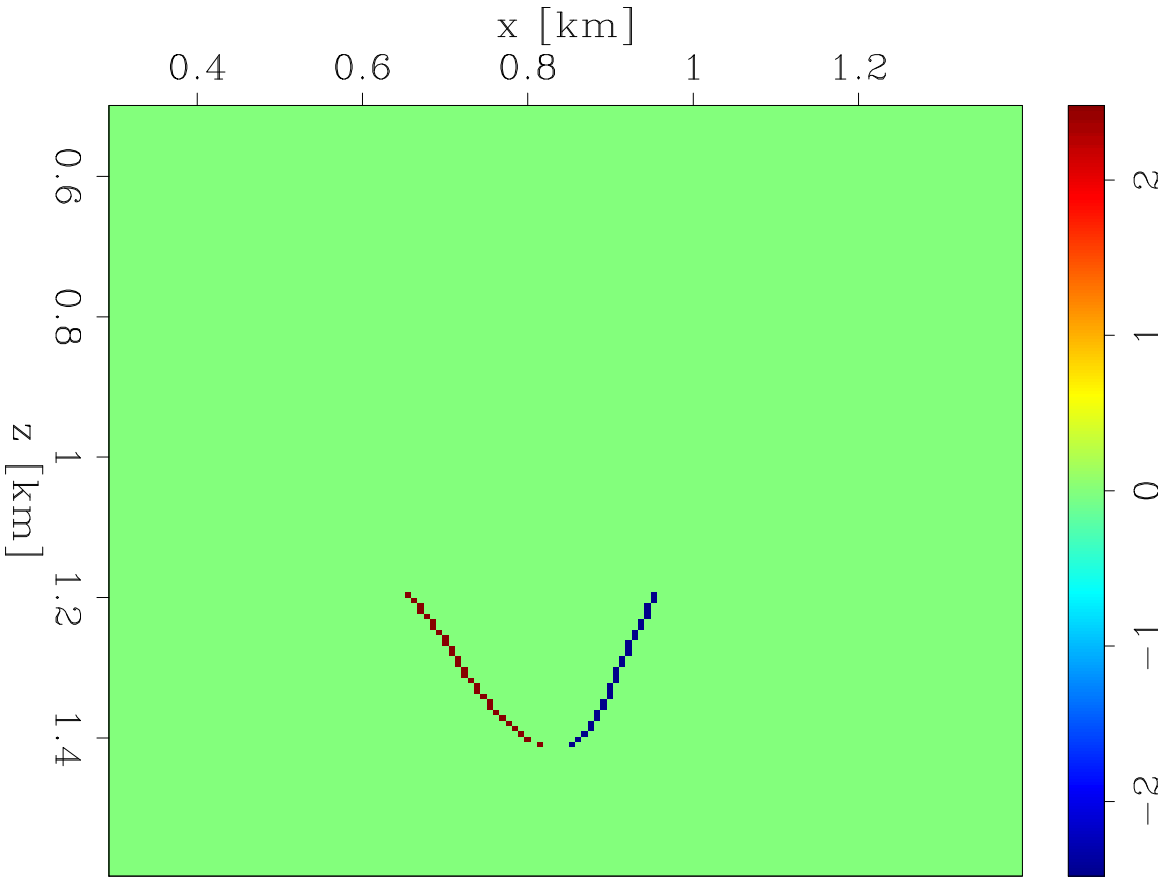


Figure 7: The double canyon perturbation of the Sigsbee model. [**ER**]

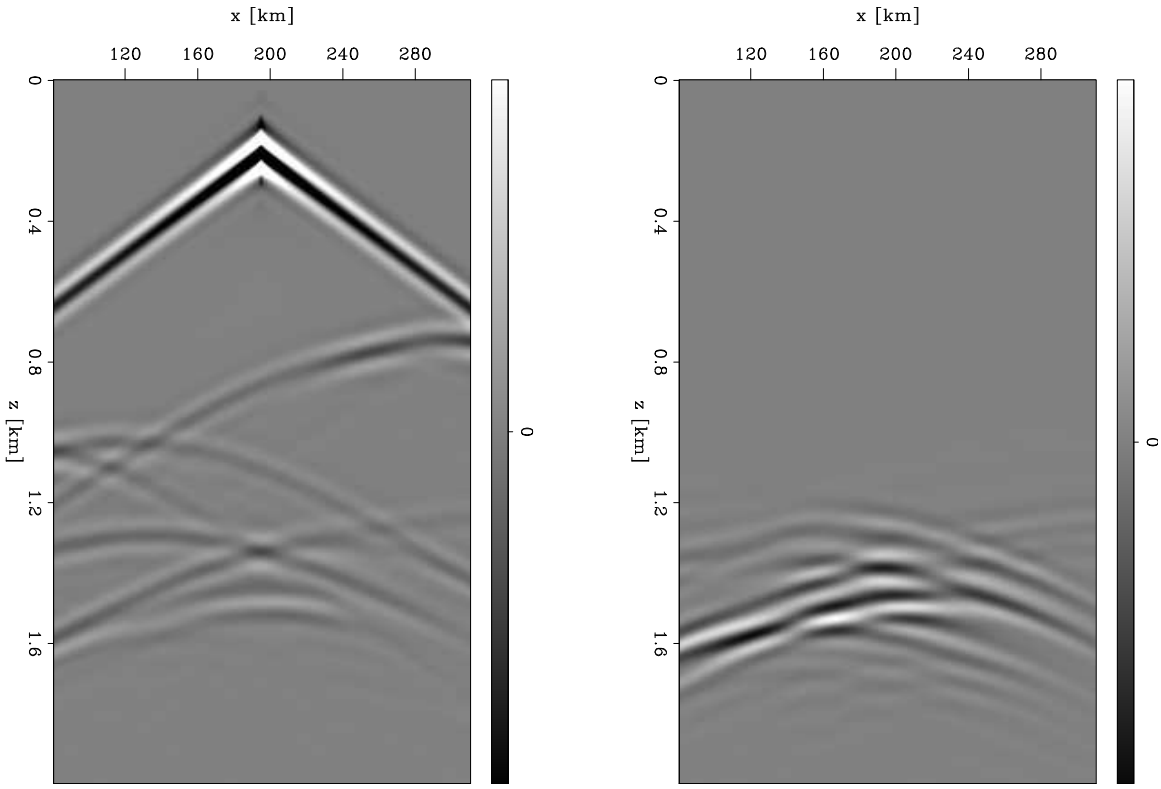


Figure 8: The data generated from the center shot on the true model (left). The center shot residual between the true data and the data generated from an initial guess that had a double canyon side perturbation (right). [**CR**]

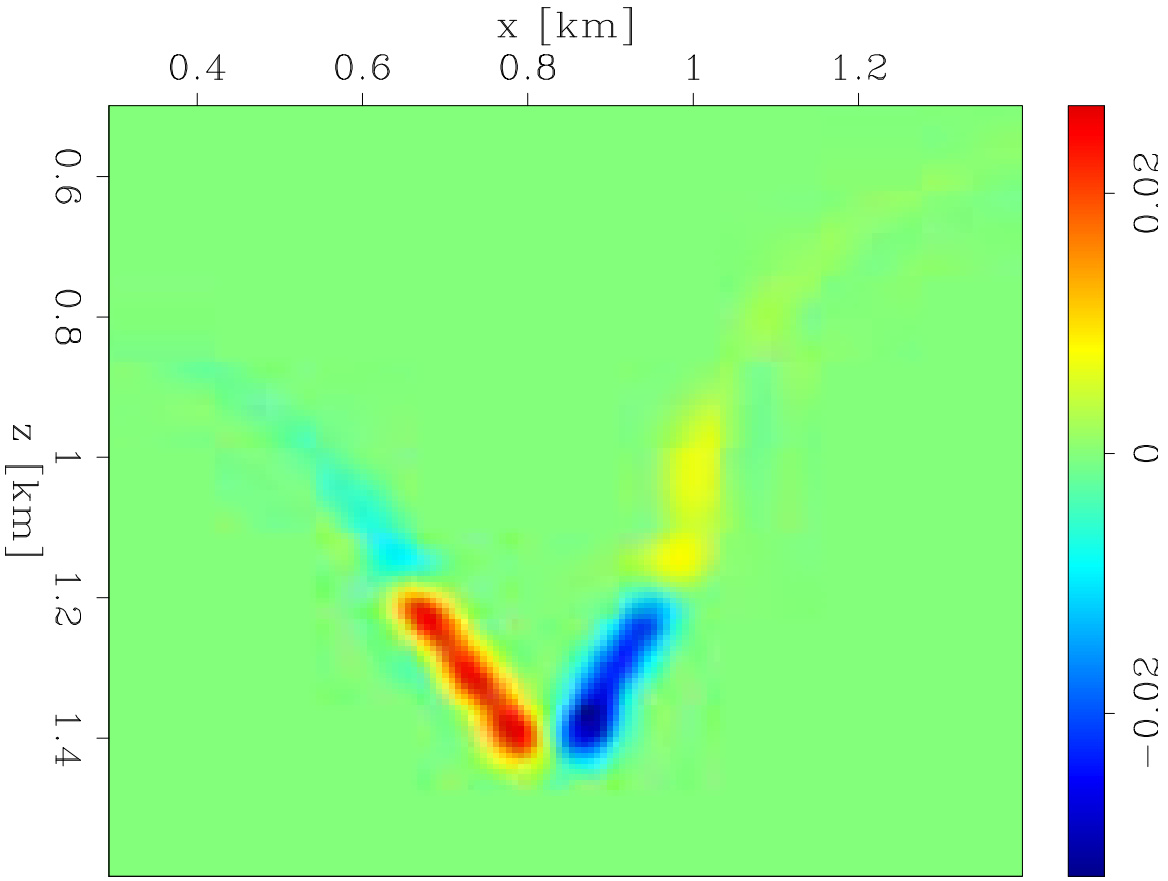


Figure 9: The inversion result after using the Gauss-Newton approximation of the Hessian on the double perturbation model. [**CR**]

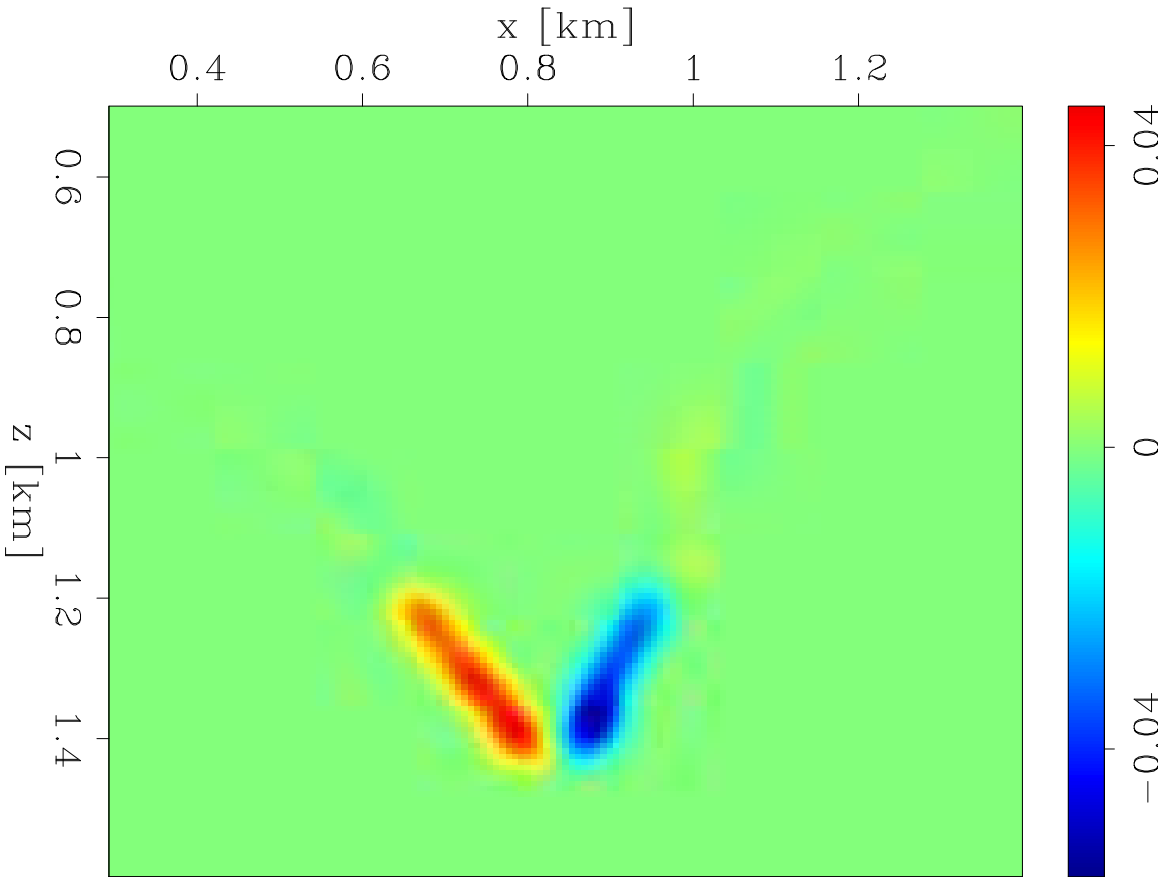


Figure 10: The inversion result after using the full Hessian on the double perturbation model. [**CR**]

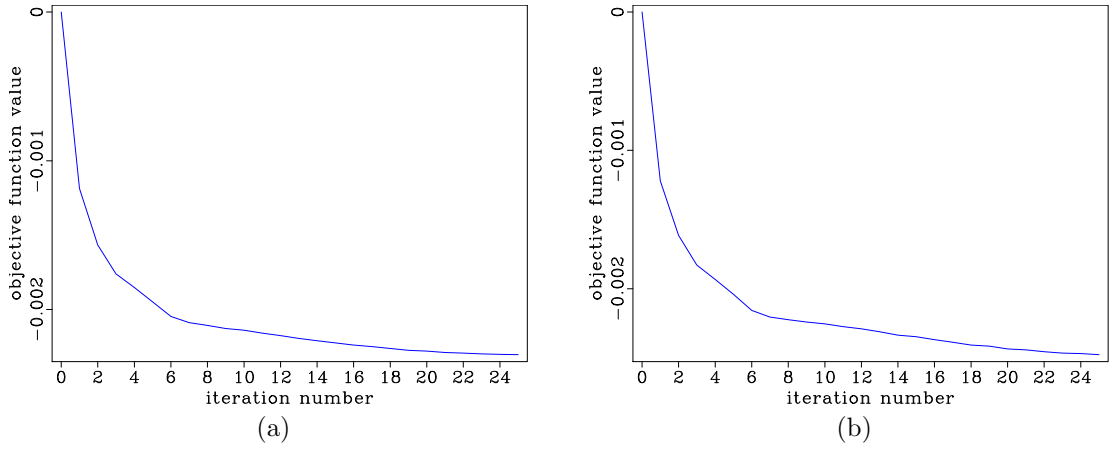


Figure 11: The objective functions from the inversions of the double canyon perturbation model using the Gauss-Newton Hessian (a), and the full Hessian (b). [**CR**]

DISCUSSION ON THE FULL HESSIAN

Benefits

The Gauss-Newton Hessian approximation is based only on the first-order scattering that the Born operator captures. This is limiting in cases where secondary scattering is more prominent, such as in salt canyons. The advantage of using the full Hessian is that its application incorporates a WEMVA term that accounts for this second order scattering. When we look at the results from the double canyon perturbation model and compare the Gauss-Newton (Figure 10) and the full Hessian results (Figure 9), we can see a slight improvement in the focusing of the energy in the full Hessian results. This improved search direction should lead to better convergence in the non-linear inversion scheme as well.

Limitations

However, we find that this improvement is not found for all models, since the Hessian is model dependent. The single perturbation example results are much different. While the Gauss-Newton Hessian system inverts quite nicely (Figures 4 and 6a), the full Hessian inversion explodes part way through (Figures 5 and 6b). Because the full Hessian operator is not inherently positive semi-definite like the Gauss-Newton Hessian is, it is possible that the operator has negative eigenvalues, which can lead to instability during inversion. This was the case in the single canyon perturbation example.

There are a number of ways this can be alleviated. One standard method is to use the Levenberg-Marquardt method Weisstein (2017) of regularizing the operator with a scaled identity matrix. However, in order to use this method properly, the correct scaling of the identity matrix must be used. If too large of a scaling is used, the operator becomes more like the identity matrix, negating the potential benefit of inverting the full Hessian system to begin with. If the scaling is too small, the system will still be ill-conditioned. The ideal scaling is slightly more than the value of the most negative eigenvalue of the operator. This makes the operator positive definite. Since our model (and as a result, our Hessian) is very large, it is impractical to store or factorize the Hessian matrix to determine the most negative eigenvalue through traditional non-iterative linear algebra methods.

Power Iteration Method

The most practical way to find the best scaling is by using the power iteration method outlined in Larson (2012) to find the maximum absolute-valued eigenvalue (positive in the case shown for Figure 12). After this has been found, we shift the operator by the negative of this value to find a new maximum absolute-valued eigenvalue.

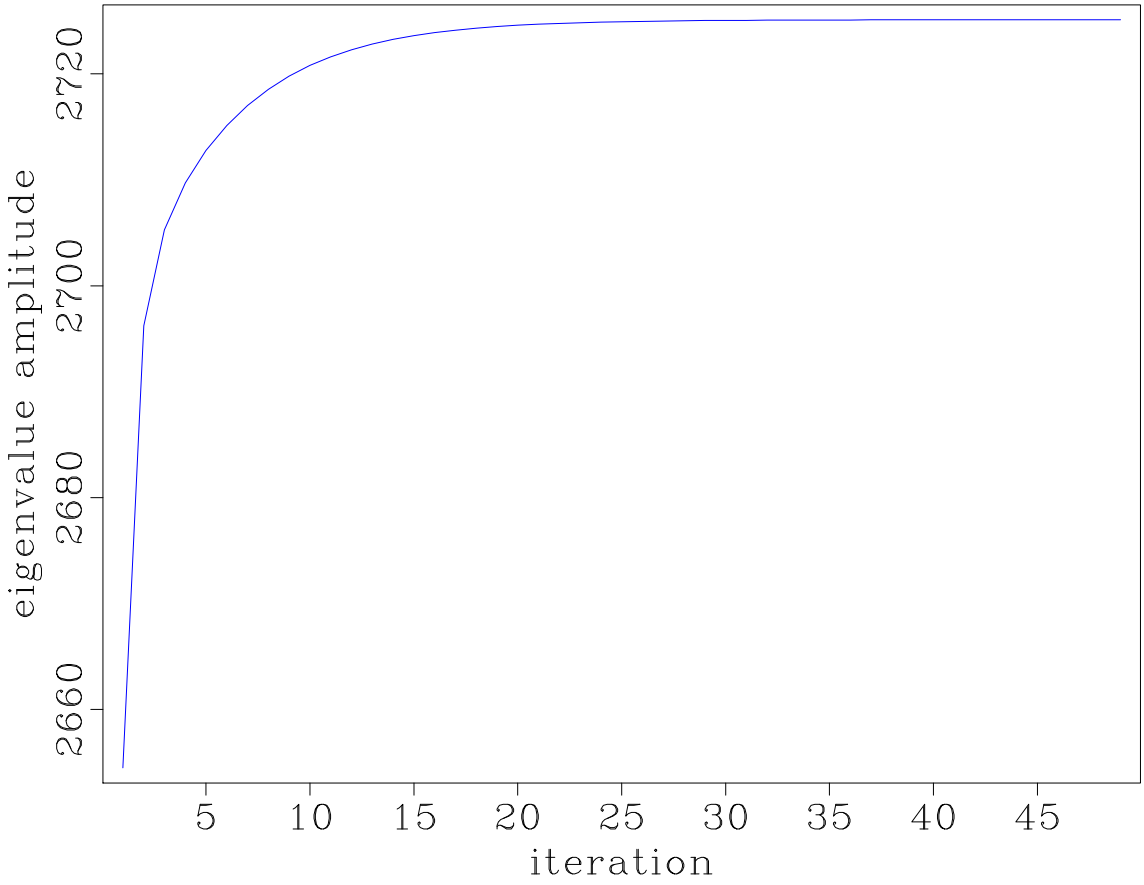


Figure 12: The power iteration curve showing the maximum absolute value approximated eigenvalues of the full Hessian operator used on the single canyon perturbation model. [**CR**]

The difference between this value and the first one derived is the magnitude of the most negative eigenvalue. We experimented with this method, but found the results of this effort to be minimal, and at notable computational cost. Figure 12 shows that in practice at least 30 iterations (and so $\sim$ 30 forward full Hessian operator applications) were necessary for each of the two power iteration searches. Once these searches were complete and a proper shift was found, we found that the results of using this Levenberg-Marquardt shift were almost imperceptible from the Gauss-Newton results. Furthermore, since the Hessian operator is model-dependent (and so changes with each outer loop iteration of FWI), we would need to perform these power iteration steps each time we were to invert the Newton system.

CONCLUSIONS

We successfully invert the Newton system for an objective function that is based on a sparse radial basis function parametrization. We find that the full Hessian formulation for the radial-basis function level sets provides marginal improvement over the Gauss-Newton Hessian for certain models where it inverts stably. However, this improvement is not necessarily worth the effort of ensuring a stable inversion result, which can be very costly for methods like Leavenburg-Marquadt combined with power iterations.

REFERENCES

Burger, M., 2003, A framework for the construction of level set methods for shape optimization and reconstruction: Interfaces and Free boundaries, **5**, 301–330.

Dahlke, T., B. Biondi, and R. Clapp, 2017a, Application of full hessian on level set formulation and parameter inversion: SEP-Report, **168**.

———, 2017b, Representing salt bodies with radial basis functions: SEP-Report, **170**.

Fichtner, A., 2010, Full seismic waveform modelling and inversion. Advances in Geophysical and Environmental Mechanics and Mathematics: Springer Berlin Heidelberg.

Guo, Z. and M. V. D. Hoop, 2013, Shape optimization and level set method in full waveform inversion with 3d body reconstruction: SEG Technical Program Expanded Abstracts, 1079–1083.

Kadu, A., T. van Leeuwen, and W. A. Mulder, 2016, Salt reconstruction in full waveform inversion with a parametric level-set method: CoRR, **abs/1610.00251**.

Larson, R., 2012, Elementary linear algebra: Cengage Learning.

Lewis, W., B. Starr, D. Vigh, et al., 2012, A level set approach to salt geometry inversion in full-waveform inversion: 2012 SEG Annual Meeting.

Li, C., C. Xu, C. Gui, and M. Fox, 2010, Distance regularized level set evolution and its application to image segmentation: Image Processing, IEEE Transactions on, **19**, 3243–3254.

Osher, S. and J. A. Sethian, 1988, Fronts propagating with curvature-dependent

speed: algorithms based on hamilton-jacobi formulations: Journal of computational physics, **79**, 12–49.

Santosa, F., 1996, A level-set approach for inverse problems involving obstacles: ESAIM Controle Optim. Calc. Var, **1**, 17–33.

Weisstein, E. W., 2017, Levenberg-marquardt method: From MathWorld–A Wolfram Web Resource.