Hessian analysis of tomographic full waveform inversion operators

Ali Almomin and Biondo Biondi

ABSTRACT

Tomographic full waveform inversion (TFWI) provides a framework to invert the seismic data that is immune to cycle-skipping problems. This is achieved by extending the wave equation and adding a spatial or temporal axis to the velocity model. For computational efficiency, the inversion is performed in a nested scheme. We examine the linearized component of the nested inversion scheme and present alternative fitting goals that have different properties compared to the original formulation. Then, we compute the Hessian matrix of both formulations as well as their individual operators to analyze the properties of each matrix. The analysis of the new formulation indicate an improved convergence behavior of inversion.

INTRODUCTION

Tomographic Full Waveform Inversion (TFWI) (Symes, 2008; Biondi and Almomin, 2012) provides a way to overcome cycle-skipping problems by combining both FWI and wave-equation migration velocity analysis (WEMVA) techniques in a generalized framework. This generalized approach utilizes the components of all seismic data to invert for the medium parameters without cycle-skipping. This is achieved in two steps: first, extending the wave equation and adding a subsurface offset axis to the velocity model, and second, adding a regularization term that drives the solution towards the zero subsurface offset. Biondi and Almomin (2013b) presented an alternative extension using time shift instead of subsurface offset. In either setting, this velocity extension makes the propagation considerably more expensive because each multiplication by velocity becomes a convolution over the extended axis.

In a companion abstract (Biondi and Almomin, 2013a), we presented an approximation that significantly reduces the computational cost of TFWI by breaking the velocity model into a background component and a perturbation component. We achieve this in two steps. First, we set up a nested inversion scheme that utilizes the nonlinear modeling operator to update the residuals. Second, the two components of the gradient are first mixed and then separated based on a Fourier domain scale separation.

In this report, we examine the properties of the inner loop of the nested scheme. Then, we present alternative fitting goals that have different properties compared to

the original formulation. Finally, we perform a Hessian analysis on both fitting goals and their operators in order to evaluate each formulation.

FITTING GOALS

We start with the full waveform objective function, J_{FWI}, which we write as

$$J_{\text{FWI}}(\mathbf{s}^2) = \|\mathcal{L}(\mathbf{s}^2) - \mathbf{d}_{\text{obs}}\|_2^2, \tag{1}$$

where $\mathbf{s}$ is the slowness model, $\mathcal{L}(\mathbf{s}^2)$ is the wave-equation modeling operator, and $\mathbf{d}_{\text{obs}}$ is the observed surface data. This first objective function represents the outer loop of the inversion. We now compute the nonlinear residuals $\Delta\mathbf{d}$ as:

$$\Delta\mathbf{d}(\mathbf{s}^2) = \mathcal{L}(\mathbf{s}^2) - \mathbf{d}_{\text{obs}}. \tag{2}$$

The nonlinear residual will be used as the "observed" data for the inner loop. In the inner loop of the inversion, we need to separate the slowness model into a background and a perturbation as follows:

$$\mathbf{s}^2 = \mathbf{b}^2 + \mathbf{p}^2(\tau), \tag{3}$$

where τ is the extension axis with time lags, $\mathbf{b}^2$ is the background component, which is a smooth version of the slowness squared and $\mathbf{p}^2(\tau)$ is the perturbation component. The perturbation component can extend across several subsurface offsets or time shifts so it is important to keep its extended axis. On the other hand, the background component is not expected to generate reflections that would be grossly time shifted with respect to the recorded data, and it thus safe to reduce its extension. This greatly reduces our cost since the convolution with slowness in propagation becomes a multiplication. The model separation allows us to use a linearized (Born) operator $\widetilde{\mathbf{L}}$ to model "linearized" data. The linearized objective function J_{LTFWI} can be written as:

$$J_{\text{LTFWI}}(\mathbf{b}^2, \mathbf{p}^2(\tau)) = \|\widetilde{\mathbf{L}}(\mathbf{b}^2)\mathbf{p}^2(\tau) - \Delta\mathbf{d}\|_2^2 + \|\epsilon\mathbf{A}\mathbf{p}^2(\tau)\|_2^2, \tag{4}$$

where $\widetilde{\mathbf{L}}$ is the Born modeling operator, $\mathbf{A}$ is the regularization operator that attempts to focus the extended model $\mathbf{p}^2(\tau)$ and ϵ is a scalar to balance the two terms of the objective function. The Born modeling operator is linear with respect to perturbation but nonlinear with respect to the background component. Hence, another linearization around the current value for both model parameters is required to compute the gradient. This is achieved by Taylor's expansion of the operator around the model estimates at the current inner loop index i, i.e. $\mathbf{b}_i^2$ and $\mathbf{p}_i^2(\tau)$, and dropping the higher-order terms as follows:

$$\begin{aligned}\widetilde{\mathbf{L}}(\mathbf{b}_i^2 + \Delta\mathbf{b}^2)\left[\mathbf{p}_i^2(\tau) + \Delta\mathbf{p}^2(\tau)\right] &\approx \widetilde{\mathbf{L}}(\mathbf{b}_i^2)\mathbf{p}_i^2(\tau) + \widetilde{\mathbf{L}}(\mathbf{b}_i^2)\Delta\mathbf{p}^2(\tau) + \frac{\partial\widetilde{\mathbf{L}}}{\partial\mathbf{b}^2}(\mathbf{b}_i^2)\left[\mathbf{p}_i^2(\tau)\right]\Delta\mathbf{b}^2 \\ &= \widetilde{\mathbf{L}}(\mathbf{b}_i^2)\mathbf{p}_i^2(\tau) + \widetilde{\mathbf{L}}(\mathbf{b}_i^2)\Delta\mathbf{p}^2(\tau) + \mathbf{T}(\mathbf{b}_i^2)[\mathbf{p}_i^2(\tau)]\Delta\mathbf{b}^2.\end{aligned} \tag{5}$$

The tomographic operator $\mathbf{T}[\mathbf{p}^2(\tau)]$ correlates a background and a scattered wavefield from both the source and receiver sides. The scattered wavefields are computed by correlating a background wavefield with the perturbation model $\mathbf{p}_i^2$ and then propagating again to all model locations. This operator is similar to the WEMVA operator except that in the WEMVA operator, we keep the data fixed and vary the image while here it is the other around for this tomographic operator.

We can now define the linearized form $of J_{\text{LTFWI}}$ as follows:

$$\begin{aligned} J_{\text{LLTFWI}}(\Delta\mathbf{b}^2, \Delta\mathbf{p}^2(\tau)) =& \|\widetilde{\mathbf{L}}(\mathbf{b}_i^2)\Delta\mathbf{p}^2(\tau) + \mathbf{T}(\mathbf{b}_i^2)[\mathbf{p}_i^2(\tau)]\Delta\mathbf{b}^2 - \Big(\Delta\mathbf{d} - \widetilde{\mathbf{L}}(\mathbf{b}_i^2)\mathbf{p}_i^2(\tau)\Big)\|_2^2 \\ &+ \|\epsilon\mathbf{A}\Delta\mathbf{p}^2(\tau) - \big(-\epsilon\mathbf{A}\mathbf{p}_i^2(\tau)\big)\|_2^2, \end{aligned} \tag{6}$$

where the fitting goals can be written in matrix notation as follows:

$$\begin{bmatrix} \Delta\mathbf{d} - \widetilde{\mathbf{L}}(\mathbf{b}_i^2)\mathbf{p}_i^2(\tau) \\ -\epsilon\mathbf{A}\mathbf{p}_i^2(\tau) \end{bmatrix} \approx \begin{bmatrix} \widetilde{\mathbf{L}}(\mathbf{b}_i^2) & \mathbf{T}(\mathbf{b}_i^2)[\mathbf{p}_i^2(\tau)] \\ \epsilon\mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{p}^2(\tau) \\ \Delta\mathbf{b}^2 \end{bmatrix}. \tag{7}$$

To summaries, this scheme has three levels: an outer loop that tries to fit the data with the nonlinear operator (equation 1), an inner loop that tries to fit the nonlinear residual with the Born operator (equation 4), and finally a linearized loop that computes the updates for each model component (equation 6). When we start the inner loop, we set $\mathbf{b}^2 = \mathbf{s}^2$ and $\mathbf{p}^2 = 0$. Once we iterate enough times within the inner loop, we update the velocity model with both the background and the perturbation at $\tau = 0$. Appendix A describes the derivation of the Born and tomographic operators and how to numerically evaluate them and their adjoints.

There is one potential issue in the previous formulation, particularly in equation 4. The Born operator will attempt to match the first-order scattering from the background while the focusing term will attempt to force the extended model into the zero lag. This means that the nonlinear residuals will be consistent with one term of the objective function J_{LTFWI} but not both, resulting in these parts of the objective function to compete with each other. This behavior could slow down the convergence of the inversion.

In order to avoid conflicting terms in the objective function, we propose to set alternative fitting goals such that the data fitting term is independent of the regularization term. One way to achieve this is to first compute a perturbation model $\widehat{\mathbf{p}^2}(\tau)$ as follows:

$$\widehat{\mathbf{p}^2}(\tau) = \alpha\widetilde{\mathbf{L}}'(\mathbf{b}_i^2)\Delta\mathbf{d}, \tag{8}$$

where $'$ denotes the adjoint and α minimizes the following objective function:

$$J_{\widehat{\mathbf{p}^2}}(\alpha) = \|\Delta\mathbf{d} - \alpha\widetilde{\mathbf{L}}(\mathbf{b}_i^2)\widetilde{\mathbf{L}}'(\mathbf{b}_i^2)\Delta\mathbf{d}\|_2^2. \tag{9}$$

Once the perturbation model is computed, the independent fitting goals can be set as follows:

$$\begin{bmatrix} \Delta\mathbf{d} - \widetilde{\mathbf{L}}(\mathbf{b}_i^2)\widehat{\mathbf{p}^2}(\tau) \\ -\epsilon\widetilde{\mathbf{L}}(\mathbf{b}_i^2)\mathbf{A}'\mathbf{A}\widehat{\mathbf{p}^2}(\tau) \end{bmatrix} \approx \begin{bmatrix} \widetilde{\mathbf{L}}(\mathbf{b}_i^2) & 0 \\ 0 & \mathbf{T}(\mathbf{b}_i^2)[\widehat{\mathbf{p}^2}(\tau)] \end{bmatrix} \begin{bmatrix} \Delta\mathbf{p}^2(\tau) \\ \Delta\mathbf{b}^2 \end{bmatrix}. \tag{10}$$

The residual in the first row is only a data fitting residual caused by the difference between the adjoint and inverse of $\widetilde{\mathbf{L}}$ whereas the residual in the second row is only a model regularization residual caused by not focused $\widehat{\mathbf{p}^2}(\tau)$. The perturbation model $\widehat{\mathbf{p}^2}(\tau)$ has a slightly different role than $\mathbf{p}_i^2(\tau)$ in the first formulation. In the first formulation, $\mathbf{p}_i^2(\tau)$ starts at zero but slowly change with the iterations of the inner loop whereas $\widehat{\mathbf{p}^2}(\tau)$ is computed as a scaled perturbation model before optimizing the fitting goals in equation 10.

HESSIAN ANALYSIS

A 1D model with a Gaussian anomaly (representing $\Delta\mathbf{b}$) and a spike (representing $\mathbf{p}$) is used for the synthetic examples. The true velocity is shown in Figure 1. The background velocity is 3 km/s and the Gaussian anomaly is located at 1 km. The perturbation spike is located at 2 km. A Ricker wavelet with a fundamental frequency of 15 Hz is used to model the data. There is one source-receiver pair at the top of the model.

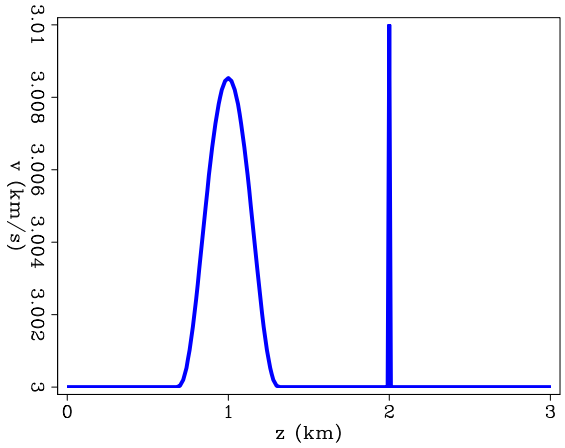

Figure 1: The true velocity model. [**ER**]

To test the $\widetilde{\mathbf{L}}$ operator, I start from a propagation velocity model that is the same as the true model except for the perturbation spike. The result of applying the adjoint of $\widetilde{\mathbf{L}}$ operator is shown in Figure 2. Next, I test the tomographic operator by computing the data residual due to removing the Gaussian anomaly. I use the results of applying the adjoint of $\widetilde{\mathbf{L}}$ as perturbation to estimate the update of the propagation velocity. The results of using the adjoint tomographic operator is shown in Figure 3. Both operators have the correct direction for the anomaly.

To compare the properties of both total operators in equations 7 and 10, we first examine the properties of their individual operators. We first compute the Hessian matrix of the $\widetilde{\mathbf{L}}$ operator at $\tau = 0$ as shown in Figure 4 and the singular values of that Hessian matrix as shown in Figure 5. The matrix is mostly diagonal with the imprint of the wavelet around the main diagonal and small artifacts due to the boundary conditions. The singular values drop significantly after approximately 100 values.

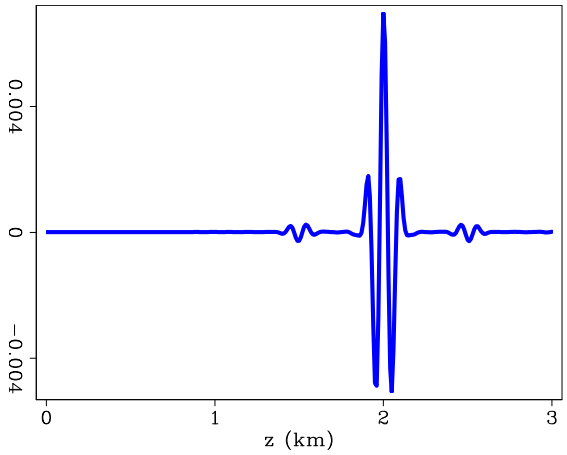

Figure 2: The Born operator adjoint on the data residual. [**ER**]

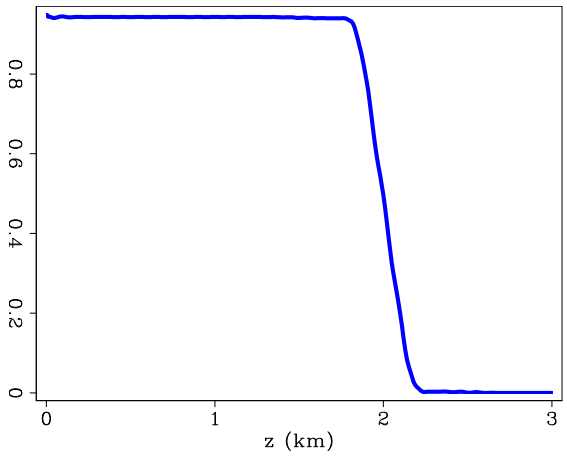

Figure 3: The tomographic adjoint on the data residual. [**ER**]

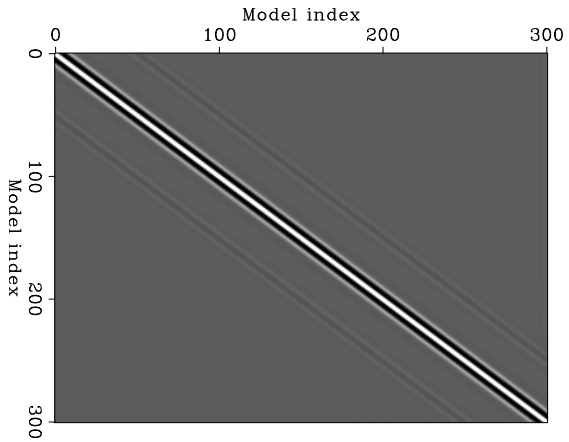

Figure 4: Hessian matrix of the $\widetilde{\mathbf{L}}$ operator at $\tau = 0$. [**ER**]

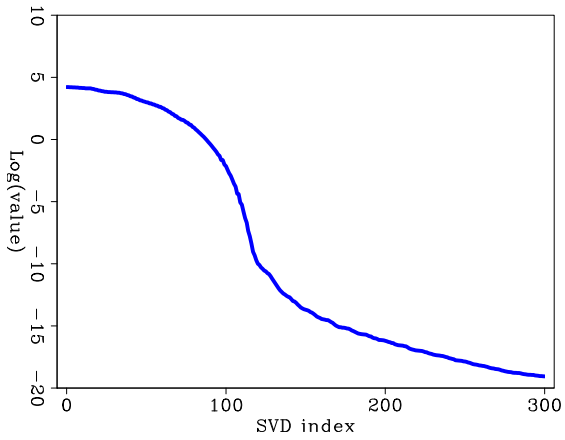

Figure 5: Singular values of Hessian matrix of the $\widetilde{\mathbf{L}}$ operator at $\tau = 0$ in log scale. [**ER**]

Next, we compute the Hessian matrix of the $\widetilde{\mathbf{L}}$ operator with τ lags from -10 to 10 samples as shown in Figure 6 and the singular values of that Hessian matrix is shown in Figure 7. The Hessian matrix is now very non-diagonal due to the convolution by the extended model. The singular values are also higher than those shown in Figure 5.

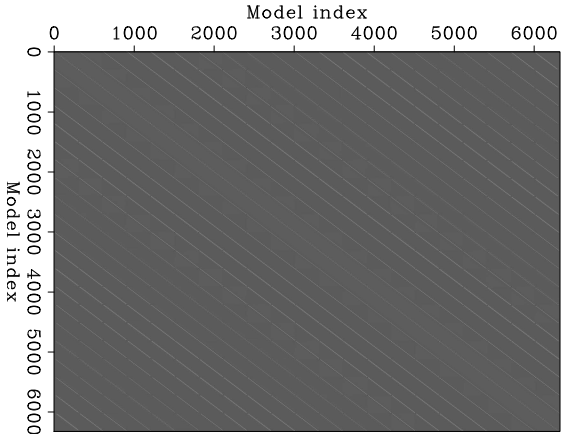

Figure 6: Hessian matrix of the $\widetilde{\mathbf{L}}$ operator. [**ER**]

The last individual operator to examine is the tomographic operator $\mathbf{T}[\mathbf{p}^2(\tau)]$. The Hessian matrix of the $\mathbf{T}[\mathbf{p}^2(\tau)]$ operator is shown in Figure 8 and the singular values of that Hessian matrix as shown in Figure 9. The hessian matrix is much less diagonal compared to the $\widetilde{\mathbf{L}}$ operator due to the integral nature of the tomographic operator. The plot of the singular values is very different than those of the $\widetilde{\mathbf{L}}$. The singular values for the $\mathbf{T}[\mathbf{p}^2(\tau)]$ operator have a relatively larger first singular value and the following values drop at a steeper rate.

Now, we compute the Hessian matrix of the total in Equation 7 and its singular values as shown in Figure 10. The singular values of this operator drop very slowly and become almost constant after 100 values. This is largely due to having $\widetilde{\mathbf{L}}$ at the off-diagonal of the total operator. Having that many large singular values indicate slow convergence rate of the inversion.

Finally, we compute the Hessian matrix of the total in Equation 10 and its singular

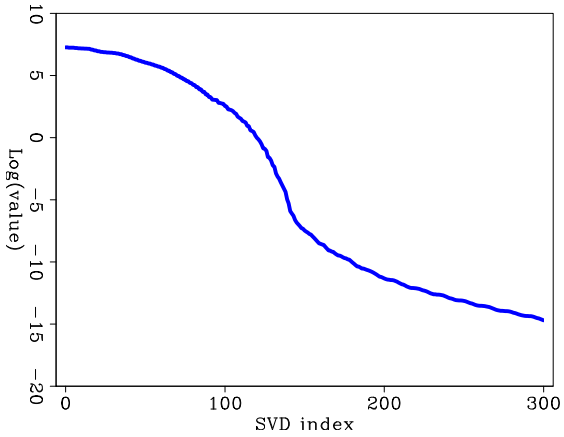

Figure 7: Singular values of Hessian matrix of the $\widetilde{\mathbf{L}}$ operator in log scale. [**ER**]

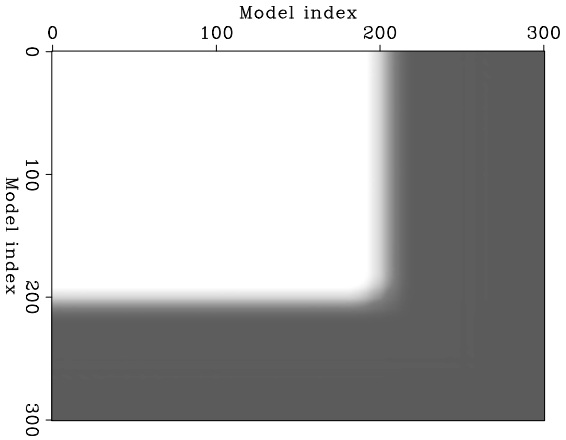

Figure 8: Hessian matrix of the $\mathbf{T}[\mathbf{p}^2(\tau)]$ operator. [**ER**]

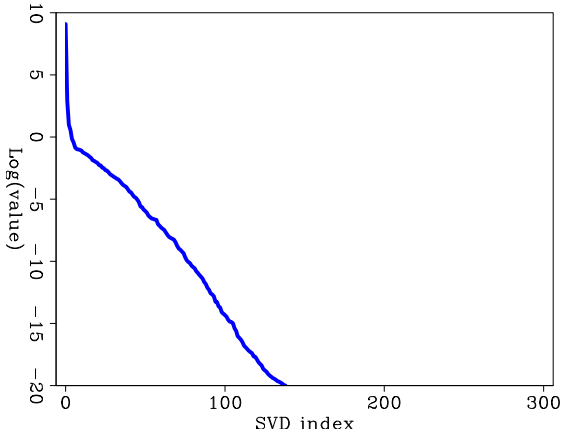

Figure 9: Singular values of Hessian matrix of the $\mathbf{T}[\mathbf{p}^2(\tau)]$ operator in log scale. [**ER**]

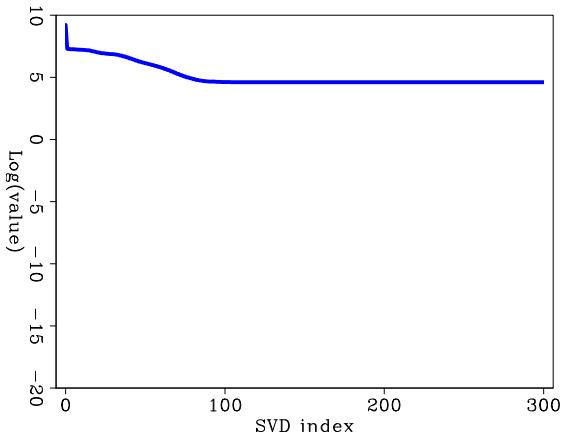

Figure 10: Singular values of Hessian matrix of the total operator in Equation 7 in log scale. [**ER**]

values as shown in Figure 11. By comparing the these singular values to those in Figure 10, we see a large decrease in the values. Moreover, the singular values of the new operator drop at a much steeper rate. These changes indicate that inverting the new total operator will have better convergence rate and properties.

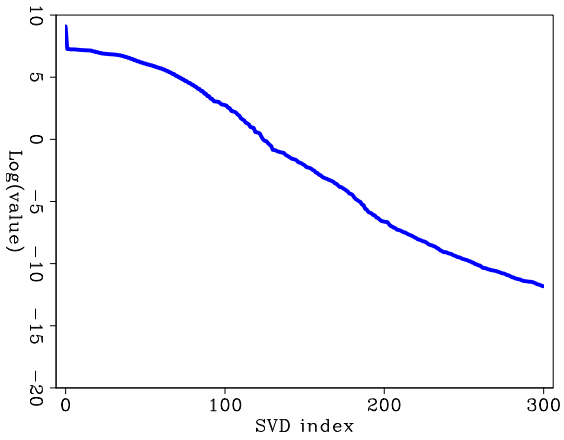

Figure 11: Singular values of Hessian matrix of the total operator in Equation 10 in log scale. [**ER**]

CONCLUSIONS

Tomographic full waveform inversion (TFWI) requires separating the slowness model into background and perturbation in order to utilize an extended model efficiently. However, a direct minimization of the resulting objective function can potentially have bad convergence properties due to the conflicting terms in the objective function. This is also verified by computing the Hessian matrix of the total operator and its singular values. We proposed an alternative way to minimize the objective function of TFWI without conflicting terms in the objective function by separating the residuals into independent data-fitting residual and model regularization residual. The singular values of the new total operator indicate an improved convergence behavior of inversion. Further testing of both operator is needed to confirm these findings.

APPENDIX A

The wave-equation modeling operator $\mathcal{L}(\mathbf{s}^2)$ is evaluated by solving the equation

$$\left[\mathbf{s}^2\mathbf{D_2} - \nabla^2\right]\mathbf{P} = \mathbf{f}, \tag{A-1}$$

where $\mathbf{P}$ is the pressure field, $\mathbf{D_2}$ is a finite-difference representation of the second derivative in time, ∇^2 is a finite-difference representation of the Laplacian, and $\mathbf{f}$ is the source function. The minimization of the objective function 4 requires the computation of the gradients with respect to both $\mathbf{b}$ and $\mathbf{p}$. These gradients can be computed by a perturbation analysis of the modeling operator $\mathcal{L}(\mathbf{s}^2)$. Rewriting the perturbed modeling equation in terms of $\mathbf{b}$ and $\mathbf{p}$, and the incident wavefield, $\mathbf{P_i}$ and the scattered wavefield, $\mathbf{P_s}$, we obtain the Born modeling operator $\widetilde{\mathbf{L}}$ by the coupled equations

$$\begin{aligned}\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_i} &= \mathbf{f}, &\text{(A-2)}\\ \left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_s} &= \mathbf{p}_i^2(\tau) \overset{\tau}{*} \mathbf{D_2}\mathbf{P_i}, &\text{(A-3)}\end{aligned}$$

where $\overset{\tau}{*}$ denotes convolution in τ. Introducing the perturbations, $\delta\mathbf{b}^2$ and $\delta\mathbf{p}^2$, in the two model variables into equations A-2–A-3, and introducing the corresponding perturbations in the wavefields, $\delta\mathbf{P_i}$ and $\delta\mathbf{P_s}$ yields the following perturbed system:

$$\begin{aligned}\left[\left(\mathbf{b}_i^2 + \delta\mathbf{b}^2\right)\mathbf{D_2} - \nabla^2\right](\mathbf{P_i} + \delta\mathbf{P_i}) &= \mathbf{f}, &\text{(A-4)}\\ \left[\left(\mathbf{b}_i^2 + \delta\mathbf{b}^2\right)\mathbf{D_2} - \nabla^2\right](\mathbf{P_s} + \delta\mathbf{P_s}) &= \left[\mathbf{p}_i^2(\tau) + \delta\mathbf{p}^2(\tau)\right] \overset{\tau}{*} \mathbf{D_2}(\mathbf{P_i} + \delta\mathbf{P_i}). &\text{(A-5)}\end{aligned}$$

By setting $\delta\mathbf{b}^2 = 0$ in equations A-4–A-5, we derive the following system of equations:

$$\begin{aligned}\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_i} &= \mathbf{f}, &\text{(A-6)}\\ \left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\delta\mathbf{P_s} &= \delta\mathbf{p}^2(\tau) \overset{\tau}{*} \mathbf{D_2}\mathbf{P_i}, &\text{(A-7)}\end{aligned}$$

which can be used to evaluate the perturbations in the scattered wavefield $\delta\mathbf{P_s}$, and consequently in the recorded data $\delta\mathbf{d} = \mathbf{S}\delta\mathbf{P_s}$, caused by perturbations $\delta\mathbf{p}^2$.

We can derive the data-space tomographic operator, $\mathbf{T}$, which relates perturbations in the scattered wavefield, $\delta\mathbf{P_s}$, to perturbations in the background model, $\delta\mathbf{b}^2$, by setting $\delta\mathbf{p}^2 = 0$ and neglecting the higher-order terms in $\delta\mathbf{b}^2$ in equations A-4–A-5.

This tomographic operator is equal to the sum of two operators, $\mathbf{T_i}$ and $\mathbf{T_s}$. The first operator, $\mathbf{T_i}$, models perturbations in the scattered wavefield caused by perturbations in the propagation of the incident wavefield:

$$\begin{aligned}\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_i} &= \mathbf{f}, &\text{(A-8)}\\ \left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\delta\mathbf{P_i} &= \delta\mathbf{b}^2\mathbf{D_2}\mathbf{P_i}, &\text{(A-9)}\\ \left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\delta\mathbf{P_s} &= \mathbf{p}_i^2(\tau) \overset{\tau}{*} \mathbf{D_2}\delta\mathbf{P_i}, &\text{(A-10)}\end{aligned}$$

The second operator, $\mathbf{T_s}$, models perturbations in the scattered wavefield caused by perturbations in the propagation of the scattered wavefield itself:

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_i} = \mathbf{f}, \tag{A-11}$$

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_s} = \mathbf{p}_i^2(\tau) \overset{\tau}{*} \mathbf{D_2}\mathbf{P_i}, \tag{A-12}$$

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\delta\mathbf{P_s} = \delta\mathbf{b}^2\mathbf{D_2}\mathbf{P_s}, \tag{A-13}$$

Both of these tomographic operators depend nonlinearly on $\mathbf{b}_i$ and linearly on $\mathbf{p}_i^2$. They have zero output when $\mathbf{p}_i^2$ is equal to zero; that is, at the first iteration of the inner loop because we set $\mathbf{p}_o^2 = 0$ as a starting model of the inner iterations. However, as we update the linearization at each iteration, starting from the second iteration the output of $\mathbf{T}$ becomes different from zero.

REFERENCES

Biondi, B. and A. Almomin, 2012, Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis: SEG Expanded Abstracts, **31**, 275–279.

———, 2013a, Simultaneous inversion of full data bandwidth by tomographic full waveform inversion (TFWI): SEP-Report, **150**, 21–42.

———, 2013b, Tomographic full waveform inversion (TFWI) by extending the velocity model along the time-lag axis: SEP-Report, **149**, 37–50.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.