Towards improving crosstalk suppression in multiparameter FWI by decorrelating parameter classes

Mariana Lume, Scott Keating and Kris Innanen

ABSTRACT

Multiparameter FWI is commonly affected by parameter crosstalk. These effects are described/corrected by the Hessian, which also impact the shape of the objective function iso-surfaces and the convergence of the optimization algorithms. This study focuses on finding an intermediate model space where the parameter classes are decorrelated, i.e., where the Hessian is an identity matrix, to minimize crosstalk and reach an accurate minimum that could be transformed to the ρ, V_P and V_S model space. Transformation rules between model spaces were applied in a FWI workflow, using transformation matrices ($\mathbf{T}$) constructed to satisfy constraints imposed by the Hessian of the intermediate system. Overall, this FWI method produced relatively good V_S estimations, but did not overcome a reference FWI in the V_P and ρ results, since more crosstalk was introduced. However, improvements on the structure of the Hessians with respect to those from the reference inversion were brought for some areas of the model grid, which makes the main decorrelation ideas promising to minimize these coupled effects. The drawbacks were related to a localized approach to compute $\mathbf{T}$, which might need to account, in future work, for crosstalk contributions of multiple grid cells.

INTRODUCTION

In multiparameter Full waveform Inversion (FWI), inter-parameter coupled effects or crosstalk might occur (Operto et al., 2013). Crosstalk implies that different physical properties are confused in the inversion, yielding to poorly accurate results and convergence slowness (Keating and Innanen, 2019). A common strategy to mitigate these effects involve the analysis of radiation patterns, allowing to identify which patterns overlap for a range of angles, and thus for which of the involved properties the gradient update will be similar, giving insights about the leakage that will occur (Keating and Innanen (2019); Métivier et al. (2015)). Hence, FWI workflows are designed including parameterizations based on minor correlation of their scattering patterns and, in some cases, the dominant parameter class of a particular set. On the other hand, most local optimization algorithms require a Hessian, which is a block band-diagonal matrix that quantifies and corrects crosstalk in the gradients of the objective function through their off-diagonal blocks. Thus, trade-off suppression could be accounted through manipulation of the Hessian, since no parameter correlation would exist if the off-diagonal blocks were zero (Operto et al. (2013); Métivier et al. (2015)).

Moreover, Innanen (2020a,b,c,d,e) published a series of reports explaining: (1) how re-parameterizing seismic inversion problems is equivalent to performing a coordinate transform between two systems (2) how Gauss-Newton directions are parallel to Steepest Descent directions when the Hessian is the identity matrix, producing favorable convergence properties, and (3) how a model space characterized by a Hessian with that structure promises the minimization of crosstalk.

In this study, we performed a isotropic elastic Full Waveform Inversion in the frequency domain, with the main purpose of obtaining crosstalk corrected values of density (ρ), P-wave velocity (V_P) and S-wave velocity (V_S) by inverting for an intermediate set of parameters, that ideally should not contain any leakage, and then transforming the estimates back to the original model space. To achieve this, we incorporated the transformation rules and numerical procedures proposed by Innanen (2020a,b,c,d,e), i.e., the intermediate re-parameterization was found after solving linear relationships between the models ρ, V_P and V_S (baseline or reference system) and a transformation matrix. The crosstalk correction is included within this matrix, because it is constructed in a way that allows to convert a local Hessian (extracted from a fixed point) in the baseline system to an identity matrix in the intermediate system. In that sense, we aimed to understand how the selected point controls this re-parameterization, as well as what is the scope of accuracy expected.

BACKGROUND: FWI AND MODEL SPACE TRANSFORMATIONS

The most common objective function to minimize in FWI problems, while using local optimization strategies, is the L2 norm or least-squares (Tarantola, 1984). The wavefield dependent data fitting in frequency domain takes the form:

$$\phi = \sum_{n=1}^{N_f} \sum_{m=1}^{N_s} \frac{1}{2} \left|\left|\mathbf{R}\mathbf{u}_{\mathbf{nm}} - \mathbf{d}_{\mathbf{nm}}\right|\right|_2^2 \tag{1}$$

where N_f is the number of discrete frequencies, N_s is the number of sources, **d** is the measured data, **u** is the predicted wavefield and **R** is the receiver sampling matrix. In the isotropic elastic approximation of the wave propagation, the medium is characterized by the density ρ(x,z) and the Lamé parameters λ(x,z) and μ(x,z). In this scenario, the predicted data is computed through the wave equation described by Pratt (1990):

$$\omega^2 \rho u_x + \frac{\partial}{\partial x}\left[\lambda\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z}\right) + 2\mu\frac{\partial u_x}{\partial x}\right] + \frac{\partial}{\partial z}\left[\mu\left(\frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z}\right)\right] + f_x = 0 \tag{2}$$

$$\omega^2 \rho u_z + \frac{\partial}{\partial z}\left[\lambda\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z}\right) + 2\mu\frac{\partial u_z}{\partial z}\right] + \frac{\partial}{\partial x}\left[\mu\left(\frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z}\right)\right] + f_z = 0 \tag{3}$$

where ω is the angular frequency, u_x and u_z are the horizontal and vertical particle displacements, respectively, and f_x and f_z are the horizontal and vertical source terms. These equations were solved in this study using the second order centered finite difference approach for the spatial derivatives detailed by Pratt (1990) and adding perfectly matched layers (Berenger, 1994) to avoid boundary reflections. The finite difference procedure allows to write these wave equations in matrix form as:

$$\rho\omega^2\mathbf{u}+c_{11}\nabla\left(\nabla\cdot\mathbf{u}\right)-c_{44}\nabla\times\left(\nabla\times\mathbf{u}\right)+\nabla\left(c_{11}-2c_{44}\right)\left(\nabla\cdot\mathbf{u}\right)+\nabla c_{44}\left(\nabla\mathbf{u}+\nabla\mathbf{u}^T\right)+\mathbf{f}=0 \tag{4}$$

with $c_{11}=\lambda+2\mu$ and $c_{44}=\mu$. Being structured as $\mathbf{A}(\mathbf{s})\mathbf{u}-\mathbf{f}=0$; $\mathbf{A}$ is the wavefield operator or Helmholtz matrix. Although this is the root parameterization in elastic scenarios, the wave equation can adopt any other 3 elastic parameters that are somehow related to the basic parameterization. For instance, in the case of the ρ, V_P and V_S model space, these relationships are:

$$c_{11}={V_P}^2\rho \tag{5}$$

$$c_{44}={V_S}^2\rho \tag{6}$$

Moreover, Innanen (2020a,b,c,d) explains how each re-parameterization can be seen as a transformation between cartesian and oblique coordinate systems. In order to change to a different model space, the objective function must be considered as a scalar quantity, meaning that it is invariant under transformations. However, the model update is a vector expressed in their contravariant components, while the gradient and the Hessian tensors expressed in their covariant components, meaning that they do change under transformations. To map a contravariant vector from an initial system s to a new system r and backwards, some rules are necessary:

$$s^{\upsilon}=t^{\upsilon}_{\mu}r^{\mu} \tag{7}$$

$$r^{\upsilon}=\left(t^{-1}\right)^{\upsilon}_{\mu}s^{\mu} \tag{8}$$

For a set of 2 parameters $\mu=\upsilon=2$:

$$\begin{bmatrix}a^1\left(s\right)\\a^2\left(s\right)\end{bmatrix}=\begin{bmatrix}\frac{\partial s_1}{\partial r_1} & \frac{\partial s_1}{\partial r_2}\\\frac{\partial s_2}{\partial r_1} & \frac{\partial s_2}{\partial r_2}\end{bmatrix}\begin{bmatrix}a^1\left(r\right)\\a^2\left(r\right)\end{bmatrix} \tag{9}$$

$$\begin{bmatrix}a^1\left(r\right)\\a^2\left(r\right)\end{bmatrix}=\begin{bmatrix}\frac{\partial r_1}{\partial s_1} & \frac{\partial r_1}{\partial s_2}\\\frac{\partial r_2}{\partial s_1} & \frac{\partial r_2}{\partial s_2}\end{bmatrix}\begin{bmatrix}a^1\left(s\right)\\a^2\left(s\right)\end{bmatrix} \tag{10}$$

All these rules include a transformation matrix $\mathbf{T}$ (t in indicial notation) that could be constructed to satisfy certain constraints in the problem, such as producing iso-surfaces of the objective function with a particular shape in the transformed model space. Since scalar quantities do not change under transformation of the systems, the minimization of ϕ in the new system will output r models limited by a range of s_{μ} vectors that produce the same value of the objective function and, because there is only one set of parameters associated to the minimum cost, finding the minimizer of ϕ in the r model space (intermediate model space) is equivalent to find it in the s coordinate system (original model space).

Additionally, local optimization approaches require the gradient of the objective function, and this tensor as well as the model updates are affected by the selected parameterization. The gradient takes the form:

$$\frac{\partial \phi}{\partial \mathbf{s_{i,j}}} = \mathbf{R}\mathbf{u}^T \left(\frac{\partial \mathbf{A}}{\partial \mathbf{s_{i,j}}}\right)^T \mathbf{A}^{-1} \left(\mathbf{R}\mathbf{u}^* - \mathbf{d}^*\right) \tag{11}$$

$$\frac{\partial \phi}{\partial \mathbf{s_{i,j}}} = \left(\frac{\partial \mathbf{d_p}}{\partial \mathbf{s_{i,j}}}\right)^T \Delta \mathbf{d}^* \tag{12}$$

Working with c_{11}, c_{44} and ρ as the root parameterization is beneficial to compute the terms $\frac{\partial \mathbf{A}}{\partial \mathbf{s}}$ in Equation 11, using the chain rule (Eaid, 2021). On the other hand, depending on the optimization strategy, second-order information might be included with the Hessian, but most generally through a Hessian-vector product. The linear form of the Gauss-Newton approximation of the Hessian is:

$$\mathbf{H}_{(\mathbf{i,j}),(\mathbf{k,l})} = \left(\frac{\partial \mathbf{d_p}}{\partial \mathbf{s_{i,j}}}\right) \left(\frac{\partial \mathbf{d_p}}{\partial \mathbf{s_{k,l}}}\right)^* \tag{13}$$

where j and l refer to the parameter class and i and k to the position. This expression involves the correlation between the scattered wavefield that is caused by a diffractor point on the parameter $s_{i,j}$ and the scattered wavefield caused by a diffractor point on $s_{k,l}$ (Operto et al., 2013). Operto et al. (2013) and Métivier et al. (2015) illustrate how the full Hessian is organized in 3x3 blocks, each of size $(n_z \times n_x)\times(n_z \times n_x)$, with n_z being the number of samples in the vertical direction and n_x the number of samples in the horizontal direction of the model grid. When the parameter classes are not the same ($j \neq l$), i.e., off-diagonal blocks, the full Hessian describes and corrects the existing trade-off. Then, no coupled effects between parameters of different classes would exists if these values were zero.

Figure 1 demonstrates that if the values of the full Hessian are extracted at a fixed position, a local 3x3 matrix can be formed, corresponding to a point-wise Hessian and characterizing the crosstalk between parameters of different classes only at that location. Moreover, the Hessian-vector product can be written as $\nabla g_i(s)^T \Delta s$, representing the rate of change of the gradient $g_i(s)$ by the rate of change of s (model perturbation). Figure 2 illustrates that if there is a perturbation of the parameter classes at one fixed position and we want to study the change of the gradient in all locations for all parameter classes, vertical profiles accross the selected location can be extracted and, by reshaping them per block of the full Hessian, a point probes Hessian can be constructed. This Hessian is also organized in 3x3 blocks and, although it is larger than the local 3x3 Hessian, it is still very computable.

Similarly, when objective functions have ellipsoidal iso-surfaces with pronounced eccentricities and misalignments, mixing of parameter information will occur due to problems encountered by the Steepest Descent algorithm to reach the global minimum. Innanen

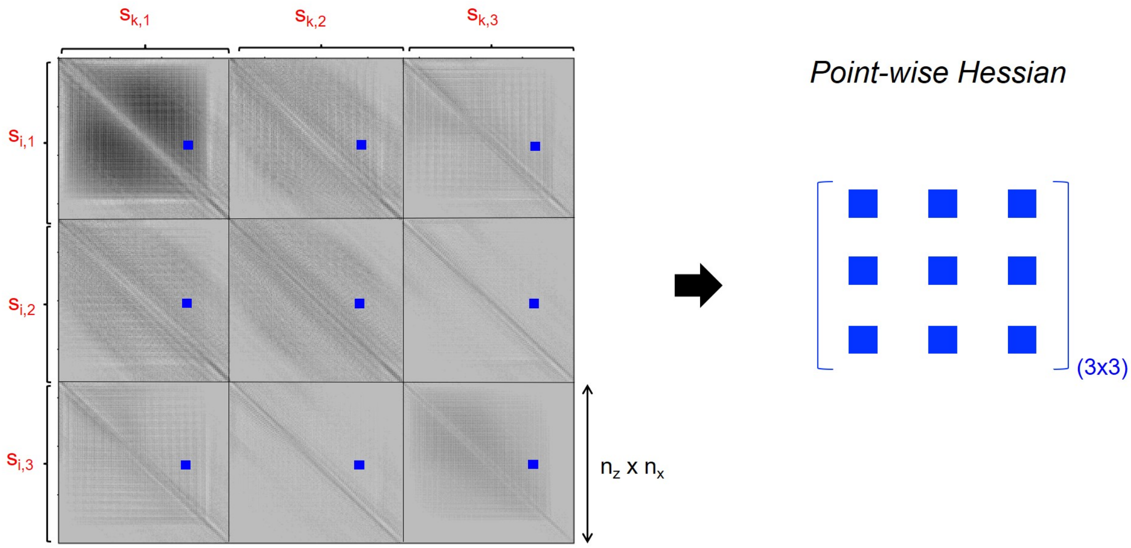

FIG. 1. Illustration of a full Hessian, modified from (Métivier et al., 2015), and how a point-wise Hessian can be constructed from it.

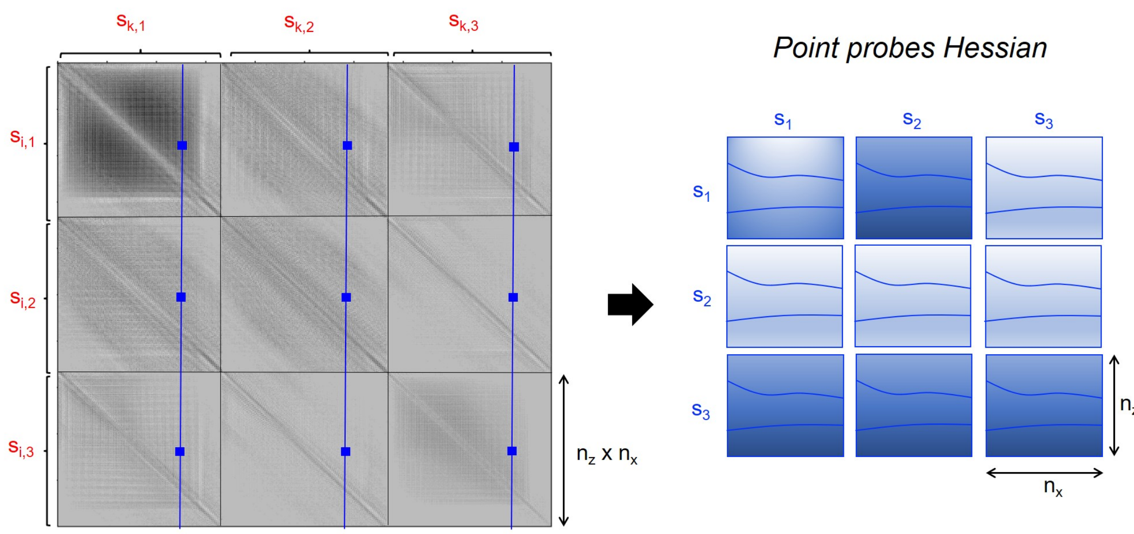

FIG. 2. Illustration of a full Hessian, modified from (Métivier et al., 2015), and how a point probes Hessian can be constructed from it.

(2020c) proved that, for a quadratic objective function, the Steepest Descent and the Gauss-Newton updates are parallel, if the Hessian behaves as an identity matrix, meaning that a more accurate local minimum can be reached with reduction of crosstalk and convergence improvement (Innanen, 2020a).

Then, we could design transformation matrices that are favorable to meet the mentioned constraints by solving the following system of equations or transformation rule for the Hessian (Innanen, 2020c):

$$t_\mu^\lambda H_{\lambda\sigma}\left(s\right)t_\upsilon^\sigma = H_{\mu\upsilon}\left(r\right) = \delta_{\mu\upsilon} \tag{14}$$

with:

$$\mathbf{T} = \begin{bmatrix} t_1^1 & t_2^1 & t_3^1 \\ t_1^{2*} & t_2^2 & t_3^2 \\ t_1^{3*} & t_2^{3*} & t_3^3 \end{bmatrix} \tag{15}$$

In a M dimensional problem for **T**, this matrix has M(M-1)/2 degrees of freedom. One way to reduce them is to pre-select and fix the lower triangular terms (marked with asterisks) and/or the diagonal entries of **T**, while computing the other entries. Innanen (2020d) explains two procedures to solve this system of equations. For this study, the first procedure was chosen, i.e., solving the system when only the lower triangular terms were fixed.

WORKFLOW AND ITS ADAPTATION FOR COMPUTATIONAL FEASIBILITY

A promising workflow entails searching for a model space where the Hessian is an identity matrix through the construction of **T** with Equation 14, and the application of the transformation rule of Equation 8, performing an efficient gradient descent strategy in this intermediate system (r model space) to find its minimum point and finally transform the result back to the original system (s model space) applying Equation 7. For this study, the Steepest Descent algorithm was preferred in order to stay the closest possible to the theoretical background. Figure 3 illustrates a scheme of the applied FWI workflow. The proposed multiscale approach was executed with 8 groups of 4 frequencies, each one starting at 1Hz and increasing its maximum value from 2Hz (first frequency band) to 20Hz (last frequency band). Additionally, the initial models were homogeneous backgrounds with the values outside the modeled heterogeneities, and 20 iterations were performed per band.

The main assumption was to work with a point-wise Hessian of the s model space and find the **T** that transforms it to an identity matrix in the r model space. This was a first attempt to make the process computationally feasible, since working with the full Hessian would imply a huge computational cost. Hence, since we would devise parameterizations based on a single point, the inversion results could be affected by the selected location to compute the transformation matrix. On the other hand, to remove degrees of freedom, the lower triangular terms in **T** were set to zero.

EVALUATION METRICS

The estimated results and the true models were compared in vertical and horizontal sections, but also with results from a FWI executed in the reference or baseline system (ρ, V_P and V_S), i.e., without applying transformation rules. Additionally, Hessians computed with the final estimates were analyzed to determine how close they were to the sought identity matrix. To achieve this, visual inspection of point probes Hessians and an appropriate evaluation metric were used. The metric corresponded to a 3x3 matrix that resulted after calculating the norm of each block of the studied point probes Hessian. It's purpose was to capture the possible existent crosstalk between parameters of different classes in all the grid cells and not only at the location selected to compute **T**. This metric indicates that the closer to zero the off-diagonal terms, the less crosstalk between parameters of different classes is encountered in all locations of the model grid. Additionally, some of the blocks

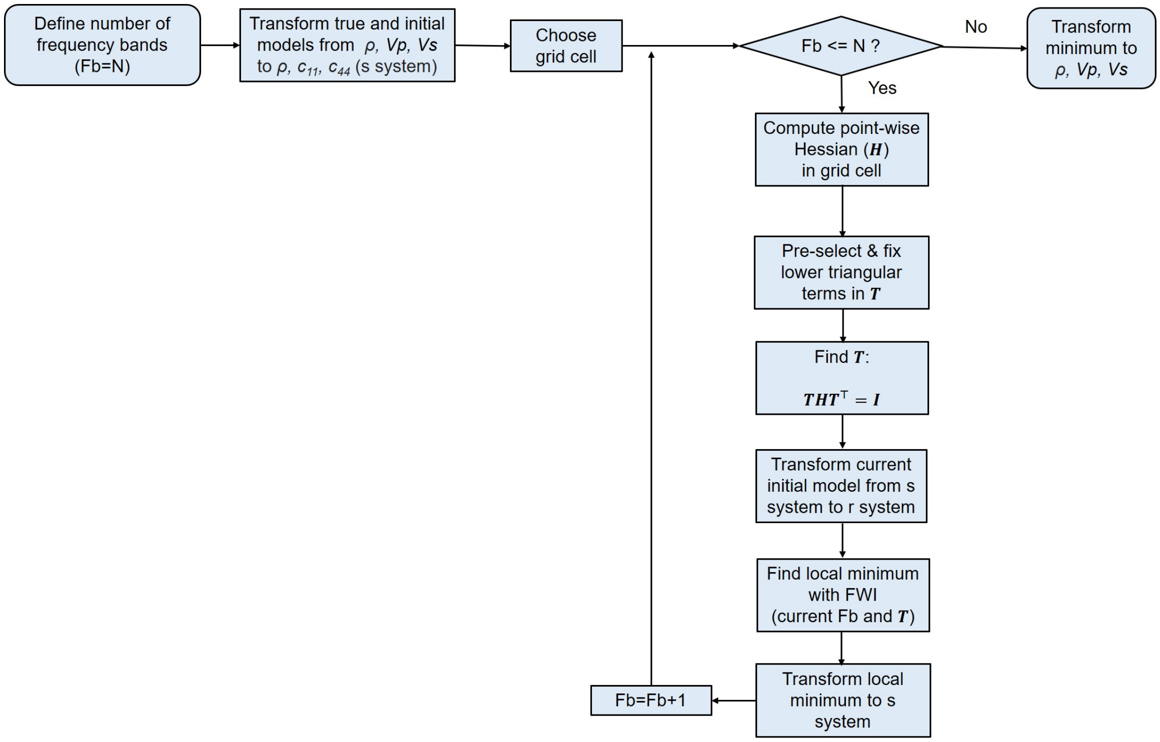

FIG. 3. Re-parameterized FWI workflow followed in this study.

of the point probes Hessian exhibited sensitivities with different orders of magnitude; thus, the crosstalk metric and local 3x3 Hessians were normalized.

RESULTS

Twenty-five sources were placed at the top of the model grid, and 98 receivers were placed at the top and bottom, to enhance illumination of the heterogeneities. Figure 4 shows the selected acquisition geometry as well as the dimension of the model grid and true values of ρ, V_P and V_S. The reference inversion was performed with the same frequency bands, initial models, optimization strategy, and number of iterations proposed for the re-parameterized approach. Figure 5 illustrates the results obtained without applying transformation rules. Noticeable crosstalk effects are observed around the ρ heterogeneity. Additionally, very subtle crosstalk is seen below the V_P anomalies.

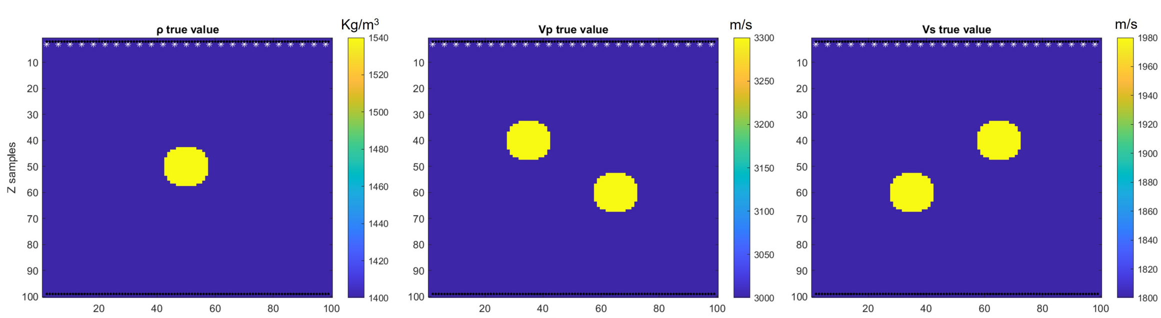

FIG. 4. True ρ, V_P and V_S models.

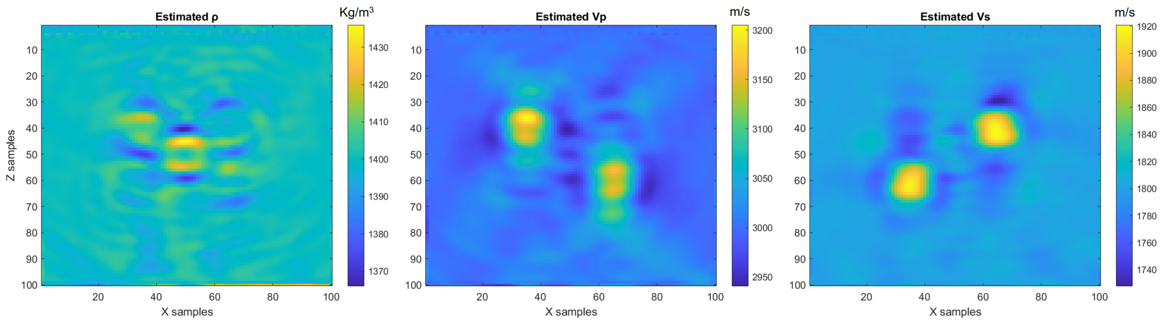

FIG. 5. Models estimated with a baseline FWI, i.e., without re-parameterization.

Figure 6 shows the point probes Hessian calculated on the estimates from the reference inversion. The point-wise Hessian computed on different grid cells are illustrated in Figure 7, while Figure 8 represents the normalized crosstalk metric after perturbing the parameter classes at different locations. Overall, at each location, the local Hessians presented a similar arrangement of values, very distant from the identity matrix. Moreover, the crosstalk metrics showed a similar organization of the values, with strong crosstalk between ρ and V_P and between ρ and V_S, but much less between V_P and V_S, in all locations, as is expected for this model space.

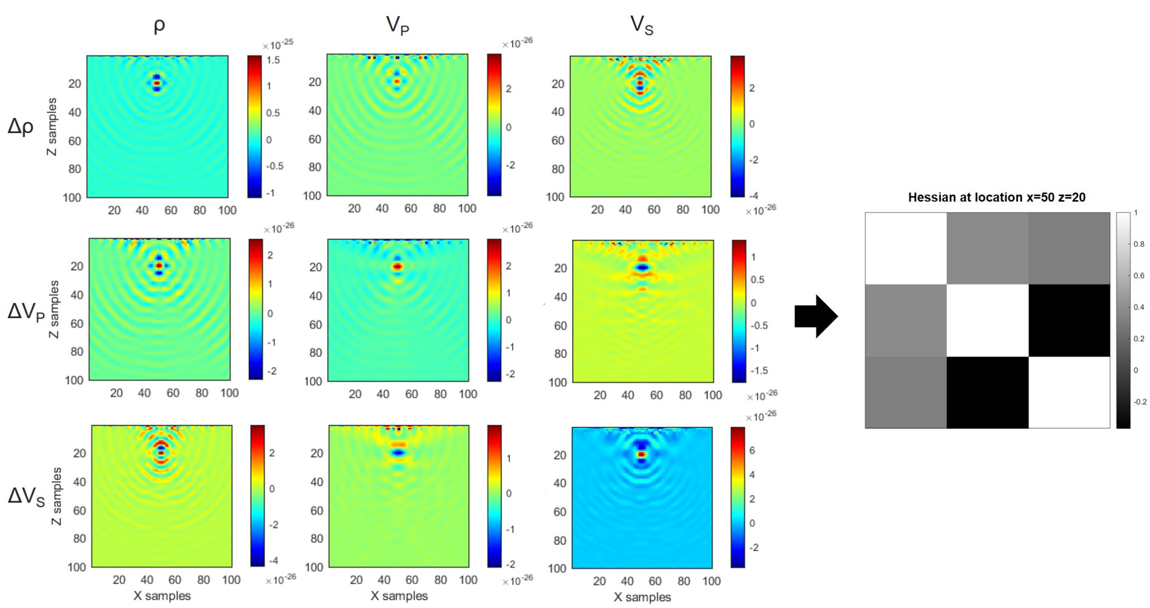

FIG. 6. Point probes Hessian computed on baseline estimates, after perturbing parameters at location x=50 and z=20. The 3x3 matrix is the local Hessian at the considered grid cell.

The models estimated with the re-parameterized FWI, choosing the location x=50 z=20 to compute $\mathbf{T}$, are shown in Figure 9. This time, much more crosstalk was encountered around the ρ heterogeneity, as well as at the top and bottom of the V_P anomalies. Figure 10 illustrates the results of the normalized crosstalk metric. There was large crosstalk between r_2 and r_1 as well as between r_3 and r_1. However, there was less crosstalk between r_2 and r_3, in all locations of the model grid. This means that V_S had much more contribution from the r_2 parameter class, while ρ and V_P had more contribution from r_1 and r_3 parameter classes, bringing their coupled effects after transforming from the r model space to the s

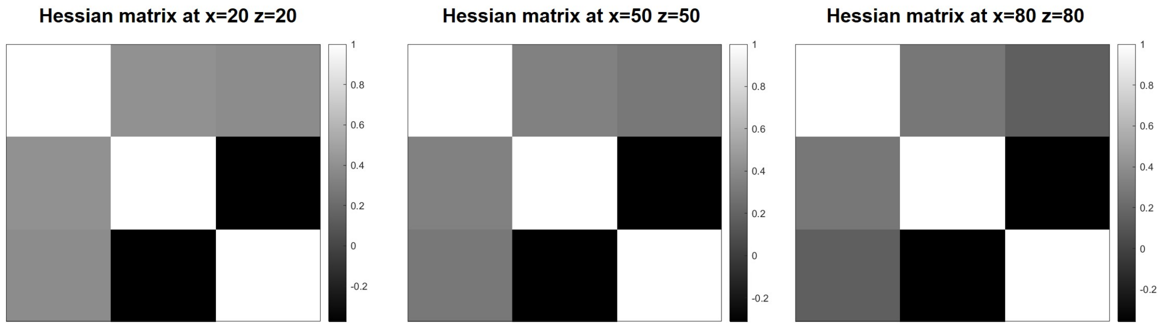

FIG. 7. Point-wise Hessians computed at different locations of the baseline estimates.

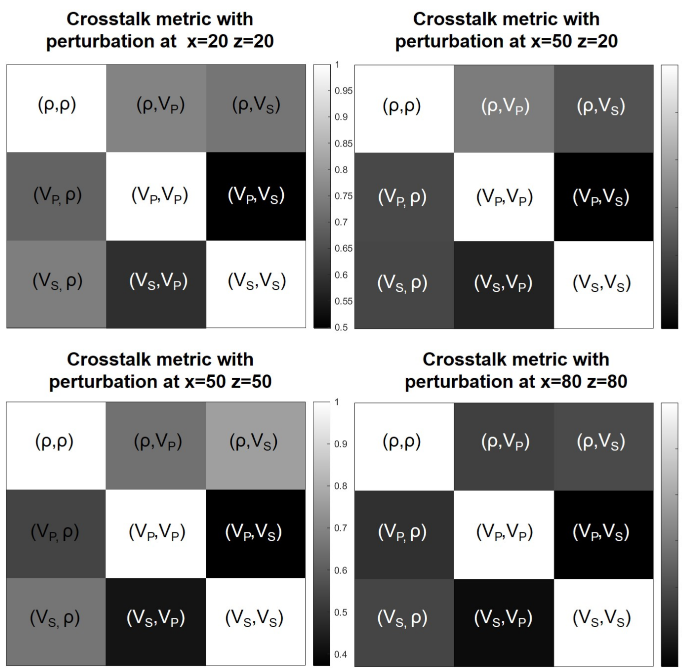

FIG. 8. Normalized crosstalk metrics computed on baseline estimates after perturbing at different locations to compute the associated point probes Hessians.

model space.

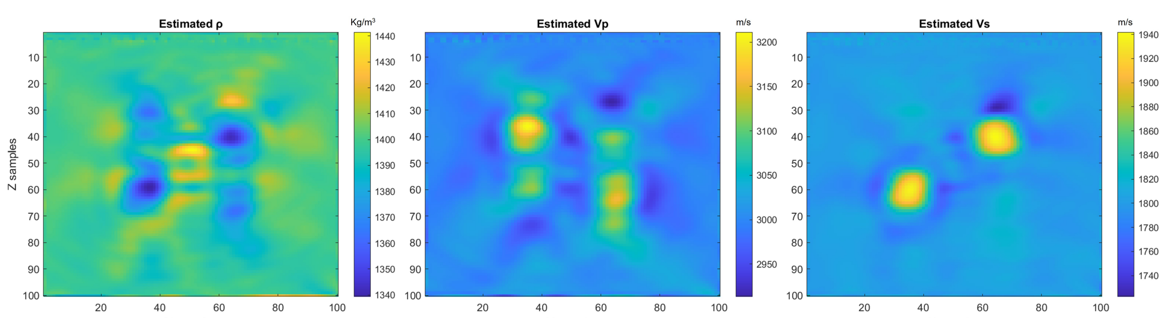

FIG. 9. Models estimated with a re-parameterized FWI after selecting location x=50 and z=20 to compute **T**.

A point probes Hessian associated to the final estimates, in the r model space, is shown in Figure 11. The identity matrix structure was observed only at and close to the grid cell

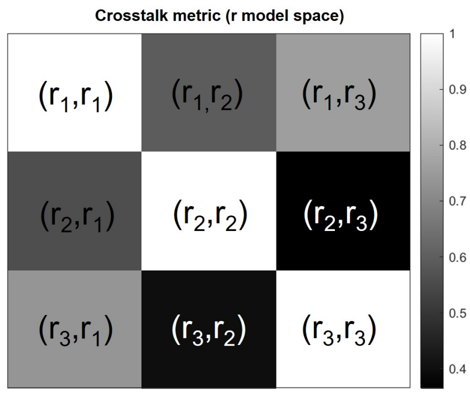

FIG. 10. Normalized crosstalk metric on estimates obtained after selecting location x=50 and z=20 to compute **T** (r model space). Parameter classes were perturbed at location x=50 and z=20 to compute the associated point probes Hessians.

chosen to compute **T**, not in the entire model grid, as would be ideally preferred to minimize crosstalk. Outside this small area, different correlation patterns appeared in the blocks of the Hessian. Hence, this explains why the crosstalk metric summarized noticeable trade-off between some model parameters in all the grid cells.

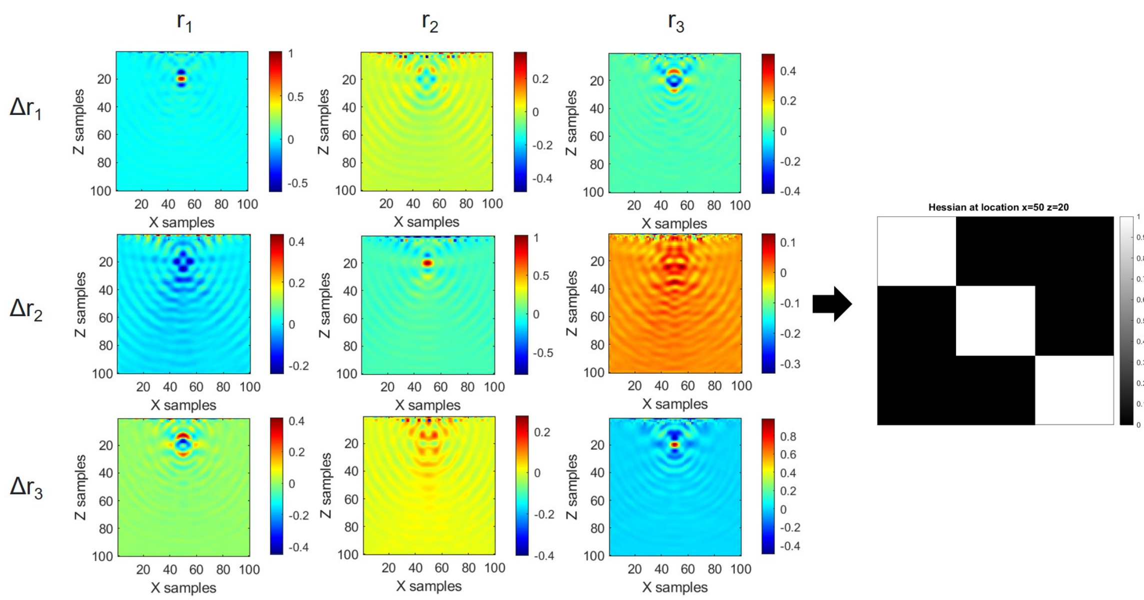

FIG. 11. Point probes Hessian computed on estimates obtained with a re-parameterized FWI after selecting location x=50 and z=20 to calculate **T** (r model space). The 3x3 matrix is the local Hessian at the considered grid cell.

Additionally, the re-parameterized FWI was performed selecting different grid cells to compute the transformation matrix. Figure 12 shows the estimated model parameters for three different locations. Moreover, Figure 13 indicates that no matter the grid cell selected, the optimization algorithm always reached the same local minimum, producing results that did not overcome the baseline inversion. Only the estimation of the V_S heterogeneity was more accurate with the re-parameterized inversion, but in terms of V_P and ρ the estimated values of the anomalies were very close to those obtained with the reference FWI and much more crosstalk was introduced rather than minimized.

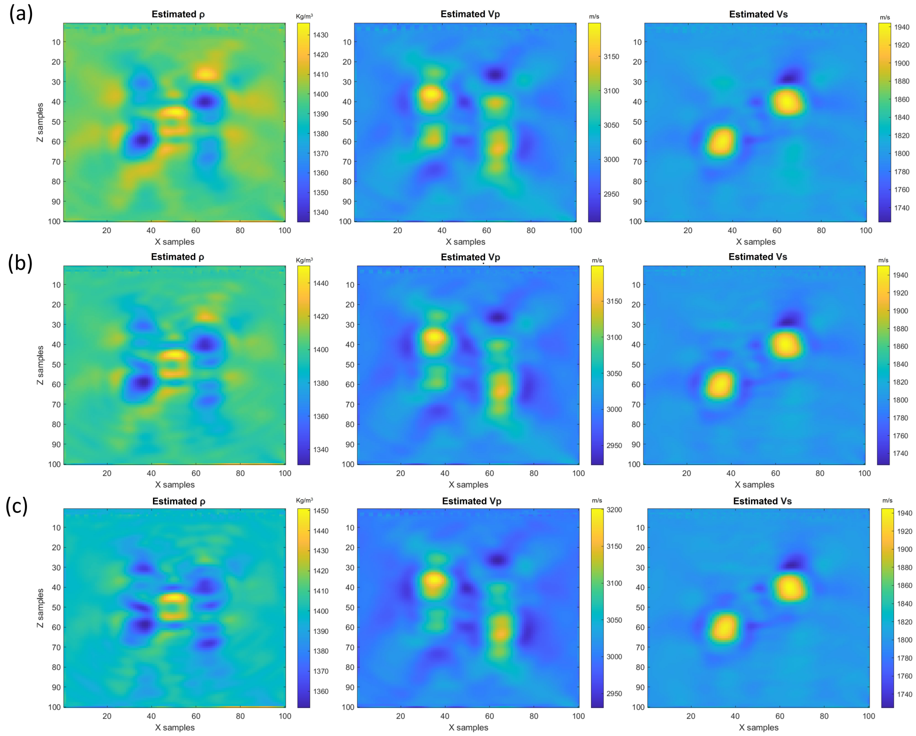

FIG. 12. Models estimated with a re-parameterized FWI using different grid locations to compute **T**: (a) x=20 and z=20, (b) x=50 and z=50, (c) x=80 and z=80.

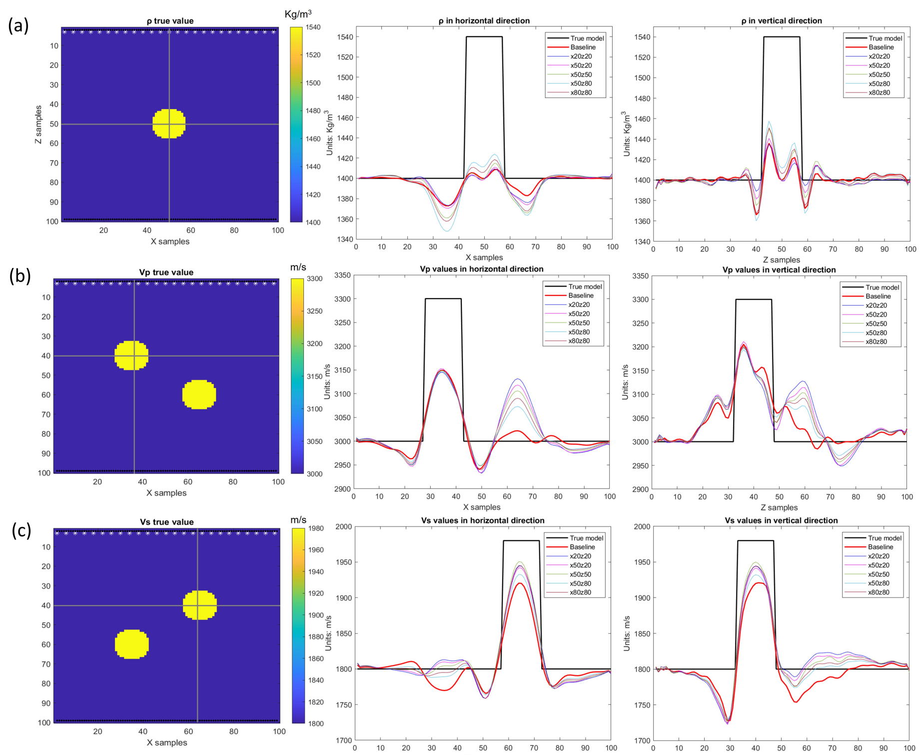

FIG. 13. Horizontal and vertical sections extracted from the models obtained with re-parameterized FWI using different pixels to compute **T**.

Figure 14 demonstrates that for grid cells (to compute **T**) close to the sources, the identity matrix was observed only around the selected pixel, but for the rest of the locations the matrix was close to the identity, yet not enough to make the iso-surfaces of the objective function spherically symmetric, since still crosstalk was produced. Conversely, as this location moves down, the identity structure was kept close to the selected pixel, but it got lost as we approached shallower depths, meaning that the objective function iso-surfaces were not circular anymore and more misalignments and eccentricities were introduced. Thus, selecting locations close to the bottom produced more crosstalk than shallower pixels.

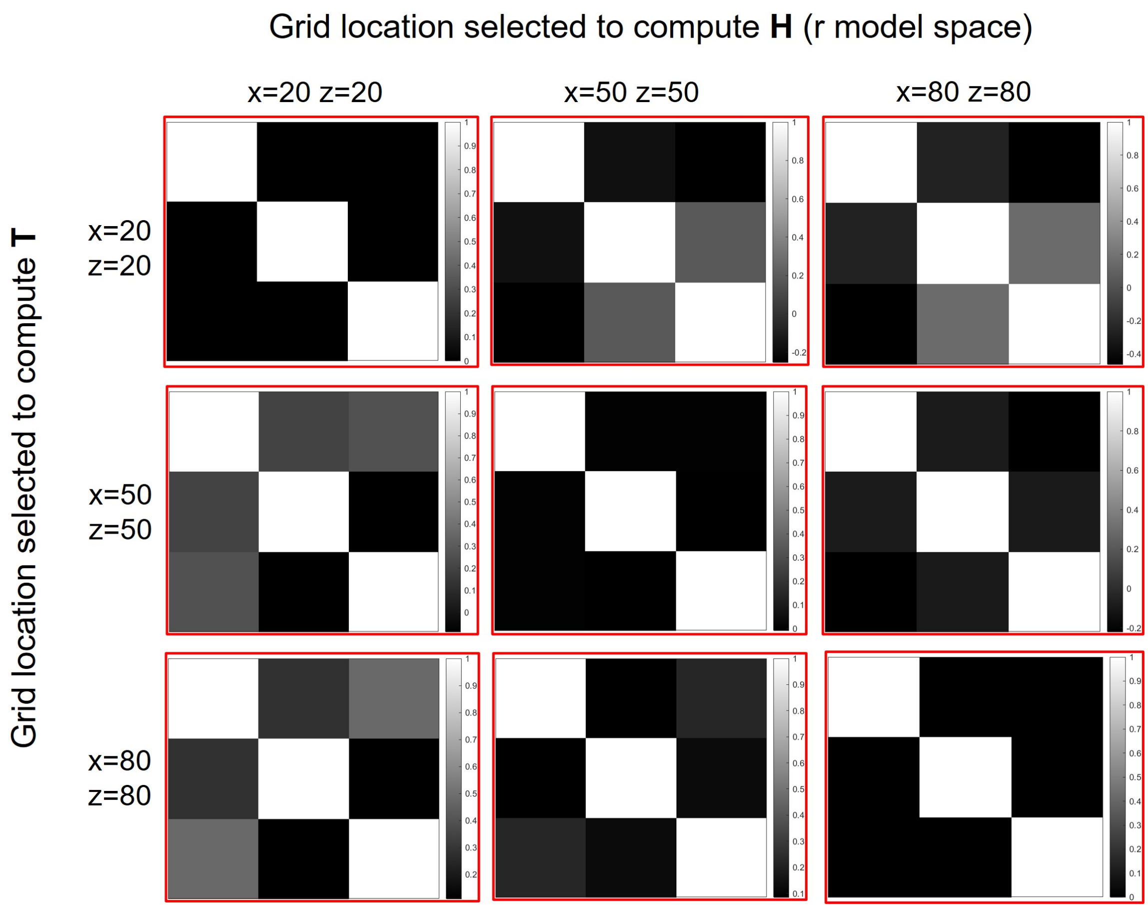

FIG. 14. Point-wise Hessian matrices at different locations of the model estimates in the r system after inverting with different pixels to compute **T**.

CONCLUSIONS

The application of the transformation rules and numerical procedures published by Innanen (2020a,b,c,d,e) into a FWI workflow did allow to find a model space where the Hessian was the identity matrix, but only in locations close to the grid cell chosen to compute the transformation matrices (**T**), generally losing this structure as we get distant from the location, exposing different crosstalk patterns outside this small area, and thus producing coupled effects between parameters of different classes. Moreover, the Steepest Descent algorithm reached the same local minimum when selecting different locations to compute **T**; hence, all models were resolved similarly, generating more accurate V_S estimates, but not better results than those from the reference inversion for V_P and ρ, since more crosstalk was introduced. In that sense, the crosstalk metric showed that the trade-off between V_P and ρ was generated by existent coupled effects between the intermediate parameter classes r_1 and r_3. Finally, although working with a transformation matrix based on a single point did not overcome the baseline inversion, the obtained results might indicate that a different numerical procedure to compute $\mathbf{T}$ should be a matter of future investigation, aiming to produce a more constant identity structure at all locations by considering the contribution of crosstalk in multiple grid cells and not only at a fixed point.

ACKNOWLEDGMENTS

The authors would like to thank the sponsors of the CREWES project as well as the NSERC (Natural Science and Engineering Research Council of Canada) under the grant CRDPJ 543578-19 for making this work possible through their financial support.

REFERENCES

Berenger, J., 1994, A perfectly matched layer for the absorption of electromagnetic waves: Journal of computational physics, **114**.

Eaid, M. V., 2021, Distributed acoustic sensing: modelling, full waveform inversion, and its use in seismic monitoring: Ph.D. thesis, University of Calgary.

Innanen, K., 2020a, Application of misfit-based model space coordinate system design to seismic AVO inversion: CREWES Research Report, **32**, No. 28.

Innanen, K., 2020b, Application of misfit-based model space coordinate system design to seismic full waveform inversion: CREWES Research Report, **32**, No. 29.

Innanen, K., 2020c, Model re-parameterization via misfit-based coordinate transforms: CREWES Research Report, **32**, No. 30.

Innanen, K., 2020d, Numerical procedures for computing constrained coordinate transforms: CREWES Research Report, **32**, No. 31.

Innanen, K., 2020e, A review of tensors in non-cartesian coordinate systems: CREWES Research Report, **32**, No. 27.

Keating, S., and Innanen, K., 2019, Parameter cross-talk and leakage between spatially-separated unknowns in viscoelastic full waveform inversion: CREWES Research Report, **31**, No. 35.

Métivier, L., Brossier, R., Operto, S., and Virieux, J., 2015, Acoustic multi-parameter FWI for the reconstruction of p-wave velocity, density and attenuation: preconditioned truncated newton approach: SEG Expanded Abstracts, 1198–1203.

Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Métivier, L., and Virieux, J., 2013, A guided tour of multiparameter Full-Waveform Inversion with multicomponent data: From theory to practice: The Leading Edge, 1040–1054.

Pratt, R. G., 1990, Frequency-domain elastic wave modeling by finite differences; a tool for crosshole seismic imaging: Geophysics, **55**.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**.