Assumptions and goals for least squares migration

Daniel Trad and Sam Gray

ABSTRACT

Least squares migration (LSMIG) uses the assumption that, if we have an operator that can create data from a reflectivity function, the optimal image will predict the actual recorded data with minimum square error. For this assumption to be true, it is also required that: a) the prediction operator must be error-free; b) model elements not seen by the operator should be constrained by other means; c) data weakly predicted by the operator should make limited contribution to the solution. Under these conditions, LSMIG has the advantage over simple migration of being able to remove interference between different model components. LSMIG does that by deconvolving or inverting the so-called Hessian operator. The Hessian is the cascade of forward modeling and migration; for each image point, it computes the effects of interference from other image points (point-spread function) given the actual recording geometry and the subsurface velocity model. Because the Hessian contains illumination information (along its diagonal), and information about the model cross-correlation produced by non-orthogonality of basis functions, its inversion produces illumination compensation and increases resolution. In addition, sampling deficiencies in the recording geometry map to the Hessian (both diagonal and non-diagonal elements), so LSMIG has the potential to remove sampling artifacts as well. These (illumination compensation, resolution, mitigating recording deficiencies) are the three main goals of LSMIG, although the first one can be achieved by cheaper techniques. To invert the Hessian, LSMIG relies on the residual errors during iterations. Iterative algorithms, like conjugate gradient and others, use the residuals to calculate the direction and amplitudes (gradient and step size), of the necessary corrections to the reflectivity function or model. Failure of conditions a), b) or c) leads the inversion to calculate incorrect model updates, which translate to noise in the final image. In this paper we will discuss these conditions for Kirchhoff migration and RTM.

INTRODUCTION

Migration is the process of decoding the subsurface geological information from seismic data. For decades this process has been done through some form of mapping from data to reflectivity. Least squares migration (Tarantola, 1984; Nemeth et al., 1999; Kuehl, 2002; Wang et al., 2003; Schuster, 2017) is a different way to decode this information, where the starting point is the operator that produces data from a reflectivity model (Figure 1). The problem then is cast as the minimization of a cost function J, which represents the global energy of the misfit or error in predicting the data properly. This goal implies a large dimensionality reduction from the dimension of the data space (number of total samples in the data) to one number, J. This simplification brings an assumption: if a reflectivity model can predict the data with minimum least squares (LS) error, then the model is "optimal". Implicitly, this statement includes other assumptions. In this paper, we will discuss these assumptions for Kirchhoff and RTM and consider ways to fulfill them with limited computer resources. Also, we will show examples for the three goals of LSMIG: illumination compensation, resolution, mitigating recording deficiencies.

NUTS AND BOLTS FOR LSMIG

Theory

Given a Born or Kirchhoff linear modeling operator $\mathbf{L}$ that can predict seismic data $\mathbf{d}$ from a reflectivity function $\mathbf{m}$ (Figure 1),

$$\mathbf{d} = \mathbf{L}\mathbf{m}. \tag{1}$$

LSMIG inverts $\mathbf{L}$ by minimizing the energy of the prediction error,

$$J = \|\mathbf{d} - \mathbf{L}\mathbf{m}\|^2. \tag{2}$$

which leads to the least squares solution:

$$\mathbf{m} = (\mathbf{L^H L})^{-1}\mathbf{L^H d}. \tag{3}$$

Here $\mathbf{L^H}$ is the adjoint operator of $\mathbf{L}$, which resembles closely the migration operator (Claerbout, 1992). The cascade action of the forward and adjoint operators, $\mathbf{L^H L}$, is known as the Hessian (Figure 2) and represents the multidimensional second derivative of the cost function with respect to the model parameters (reflectivity in this case). Also, the output of the Hessian acting on the true reflectivity model produces a blurred version of the reflectivity, or what we normally identify by migration ($\mathbf{L^H d}$).

$$\mathbf{m_{migration}} = (\mathbf{L^H L})^{-1}\mathbf{m_{true}}. \tag{4}$$

Eq. 4 is the foundation of why we identify LSMIG with the deconvolution of the Hessian operator from the migration image, and why LSMIG has higher resolution than migration. If the Hessian were stored as a matrix (which never is because of its size and complexity) then its diagonal elements would be the illumination intensities of the individual model elements. Its non-diagonal elements would the cross-correlation between different model elements (Figure 2). This interference depends not only on the shape of the basis functions (non-orthogonality), in this case Green functions, but also on its discretization (sampling) and aperture. Inverting the Hessian leads to the elimination or attenuation of the negative effects produced by the non-uniform distribution of Green functions that we use in the seismic experiment. These negative effects are illumination differences, non-orthogonality of the basis functions and sampling artifacts from both shot and receiver side.

Assumptions and goals

Although the previous discussion is straightforward and suggests that even an approximated Hessian inversion will always improve the image, in practice there are difficulties that make LSMIG a daunting process. Because migration is a linearized inversion:

(a) the prediction operator must be error-free, that is, it should have the correct mapping between data and model elements.

(b) model elements that cannot be seen by the data should be constrained by other means (regularization).

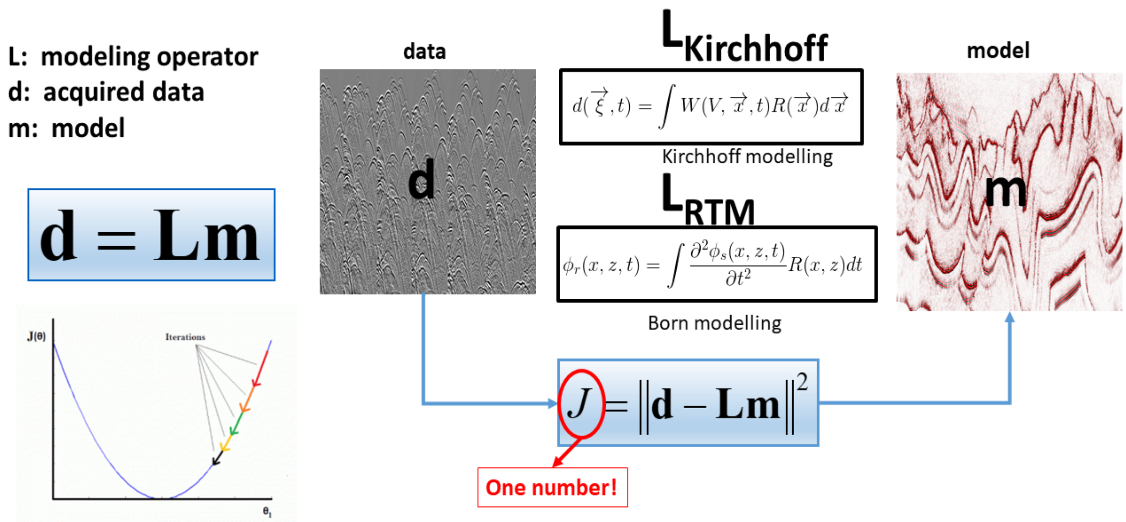

FIG. 1. Relations between data and reflectivity for Kirchhoff and RTM, when solving the direct and inverse problems

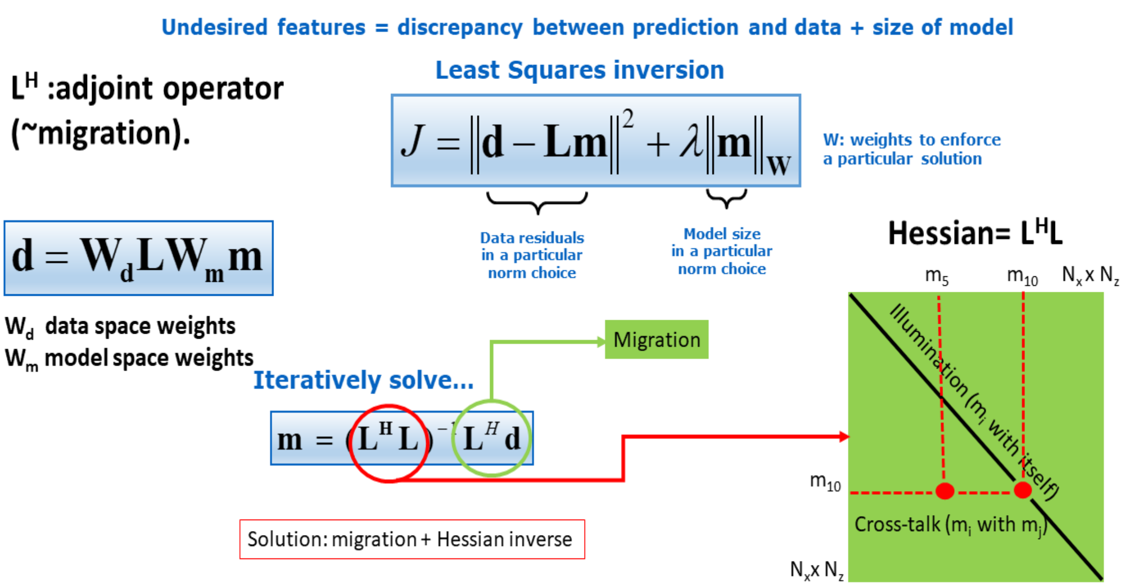

FIG. 2. LSMIG as a result of minimizing residuals and constraints. LSMIG is the deconvolution of the Hessian from the adjoint operator. The Hessian contains information about resolution and sampling

(c) data elements that are weakly predicted by the operator should make limited contribution to the solution.

To enforce these conditions, the modeling operator $\mathbf{L}$ and cost function J in Figure 1 are often modified with model and data operators $\mathbf{W_m}$ and $\mathbf{W_d}$, which can be thought of as weights or filters. These can also be interpreted as additional constraints for the cost function, like enforcing minimum model size according to energy or complexity. All these approaches are very common in optimization techniques and can be formulated also mathematically as data and model prior information by using Bayes theorem (Tarantola, 2005).

If conditions a, b, and c apply, with or without additional constraints, LSMIG has the advantage over simple mapping of being able to remove interference between different model components. As mentioned in the previous sections, LSMIG does that by deconvolving or inverting the Hessian operator (Figure 2). Because the Hessian contains illumination information (along its diagonal), and also information about the model cross-correlation produced by non-orthogonality of basis functions, its inversion produces illumination compensation and increases resolution. In addition, sampling deficiencies also map to the Hessian (both diagonal and non-diagonal elements), so LSMIG has the potential to remove sampling artifacts as well. These are the three main goals of LSMIG, although the first one (illumination compensation) can be achieved by cheaper techniques.

Illumination Compensation

Since illumination compensation is commonly applied in practice, we briefly discuss it here and show an example. The diagonal of the Hessian operator can be approximated without much computational effort by the sum of the sources auto-correlation:

$$P = \sum_{shots} \frac{\partial^2}{\partial t^2} s \bigotimes s, \tag{5}$$

where s is the source wavefield. The migration imaging condition is now normalized by this factor.

$$R = \sum_{shots} \frac{\ddot{s} \bigotimes g^*}{\sum_{shots} \ddot{s} \bigotimes s} \tag{6}$$

Here R is the reflectivity and g is the receiver wavefield. Notice that without the summation in the denominator, this term would compensate for energy differences between sources known as deconvolution imaging condition. The summation makes this compensation more general so that it also corrects for varying of shot spacing. If also had a summation for the receiver locations, then this approximation would also compensate for varying receiver intervals, but still would not contain reflector interference, which requires multidimensional deconvolution as is done in LSMIG. We illustrate this here. In Figure 3a (left) we see a blocky reflectivity model obtained by migration. In Figure 3b (centre), the imaging condition is divided by the illumination energy which compensates for an inadequate shot sampling. This compensation represents the inverse of the Hessian diagonal, and compensates for energy differences without improving resolution. This operation is commonly done in practice, since it does not involve many additional computations. Figure 3c (right) shows the effect of performing the complete Hessian deconvolution which

$$\mathbf{L^H d} \Leftrightarrow \sum_{shots}^{N} \ddot{s} \otimes g^* \qquad \mathbf{PL^H d} \Leftrightarrow \sum_{shots}^{N} \frac{\ddot{s} \otimes g^*}{\sum_{shots}(s \otimes s^*)} \qquad (\mathbf{L^H L})^{-1} \mathbf{L^H d}\,.$$

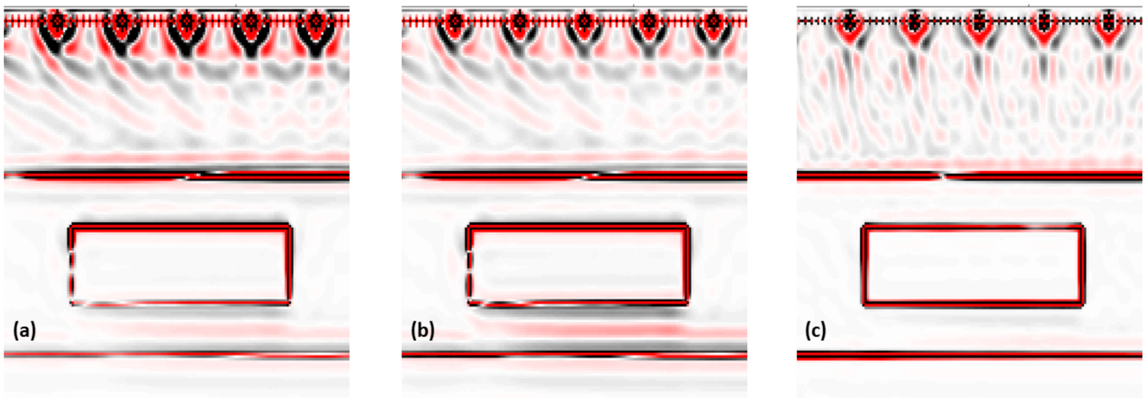

FIG. 3. Migration (left) is affected by illumination problems. Differences between shot illumination can be compensated simply by the Hessian diagonal (middle). Increasing resolution, however, requires the more expensive Hessian deconvolution (right)

increases resolution. The additional focusing of the image comes at a much larger cost than the illumination compensation.

Data and model mappings

To understand why LSMIG is more accurate than migration but less robust, depending on the accuracy of the modeling operator, we can think of the given data and the calculated model (reflectivity) as two spaces, connected through the forward and adjoint operators (Figure 4). For orthogonal operators, for example the case of well sampled Fourier transforms, the basis functions are orthogonal to each other and therefore there is no cross-correlation between different model components. Mathematically, this translates into a constant diagonal Hessian, making the adjoint and true model identical. Inversion is not needed for this case. For non-orthogonal transformations (e.g migration and demigration) model elements are connected through the non-diagonal elements of the Hessian. This is a form of cross-talk between model elements that leads to resolution loss, and depends not only on the transformation itself but also on the sampling and aperture of the physical experiment. Inversion has the potential to remove this effect, and this is the main target of LSMIG. Also, in the model space, there may be elements that do not map to the data space. For example, steep dips in the geology and under-resolved bodies may not be seen in the data. Any amount of these features can be added to the model without affecting the residuals. Simple Tikhonov regularization, or truncating the number of iterations, is in general sufficient to address this issue. Similarly, in the data there may be components that do not map to the reflectivity. An example of this is random noise. This normally does not affect the inversion (in general, adding random noise to the data is not a good way to test algorithms). The real difficulty with inversion occurs when operator $\mathbf{L}$ and its adjoint $\mathbf{L^H}$ connect the wrong elements across data and model spaces (for example because of errors in the velocity model). Although non-linear inversion has the potential to change the operator and therefore correct for this problem, it is often difficult to perform correctly. In full wave-form inversion, for example, local minima and other problems make operator corrections

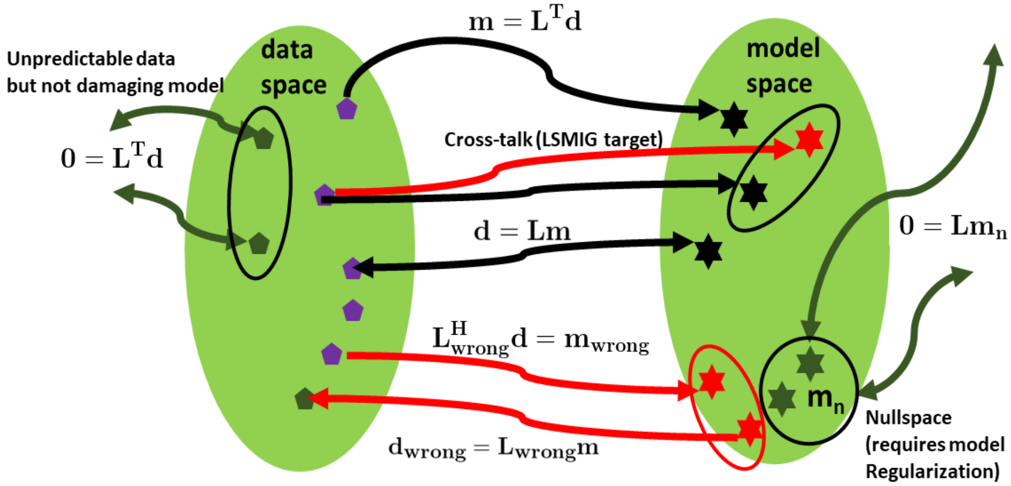

FIG. 4. Operator mapping between data and model spaces and error sources for inversion.

difficult and require more data preparation than is usually done for LSMIG.

It is possible to represent these ideas mathematically (see for example Hanke and Hansen (1993)) by using Singular Value Decomposition (SVD) for the operator:

$$\mathbf{L} = \mathbf{USV^T}. \tag{7}$$

The columns of the matrix $\mathbf{U}$ span the basis functions for the data space, and the columns of the matrix $\mathbf{V}$ (rows of $\mathbf{V^T}$) span the basis for the model. $\mathbf{S}$ is a diagonal matrix with the strength of the mappings (singular values). The solution for the LSMIG (LS inversion) can then be expressed as

$$\mathbf{m} = \sum_{i=1}^{k} \frac{\mathbf{u}_i^T \mathbf{d}}{\sigma_i} \mathbf{v}_i, \tag{8}$$

where k is the number of terms used in the inversion, less than or equal to the rank of $\mathbf{L}$, σ_i are the singular values of $\mathbf{L}$. The nullspace of the kernel $\mathbf{L}$, whose dimension is $n - p$, with n the data size, and p the rank of p, is not used in the inversion. Discarding terms larger than k has the same effect as Tikhonov regularization or stopping iterations when the residuals achieve some given threshold in iterative methods like Conjugate Gradients (CG). Eq. (8) produces correct amplitudes for $\mathbf{m}$, but it depends on a delicate balance between cross-correlation of operator and data ($\mathbf{u^T d}$) and mapping strength σ_i. If the data contain elements that the operator $\mathbf{L}$ does not produce then the cross-correlation $\mathbf{u^T d}$ can be large for very small singular values, producing arbitrarily large model artifacts . On the other hand, the result of migration (adjoint operator) is simply

$$\tilde{\mathbf{m}} = \sum_{i=1}^{k} (\mathbf{u}_i^T \mathbf{d}) \sigma_i \mathbf{v}_i, \tag{9}$$

We can see why migration, Eq. (9), produces wrong amplitudes but might be more robust than LSMIG (Eq. 8).

Numerical examples

The mathematics behind our LSMIG examples evolves around the operators $\mathbf{L}$ that map reflectivity or model $\mathbf{m}$ to the data $\mathbf{d}$, and its corresponding inversion procedure (Figure 1). We see that in both cases the reflectivity $R(x, z)$ is convolved with the wavelet after some corresponding delays to account for the wave propagation process. Usually we refer to this as Kirchhoff modeling when the directivity of the impinging wave is taken into account, or Born modeling when is not (as for RTM).

In the next figures we see examples of the three major goals of LSMIG mentioned earlier. Figure 5 shows an RTM image and Figure 6 a LSRTM for data synthesized from the Marmousi model. Illumination compensation appears as better continuity of reflectors (e.g. shallow fault blocks) in the LSRTM. Figures 7 and 8 show RTM and LSRTM for a complex Foothills model. We see a better focusing of events in the LSRTM image. Some weak reflectors are visible in the LSRTM but not in the RTM image. Figures 9 and 10 show details from Kirchhoff and LS Kirchhoff migration, for the same model. Similarly as we saw for RTM, we see an increase of resolution for LSMIG. Both images, migration and LSMIG, have deficiencies because of the difficulties for traveltime tables to exactly reproduce the traveltimes and amplitudes obtained by the finite-difference modeling. Nonetheless, the LSMIG algorithm has managed to focus the reflectors better. Using Kirchhoff operator for finite difference data is the best way to avoid the so-called inverse-crime pitfall. Finally, in Figures 11 and 12 we see the attenuation of sampling artifacts. In this case the data are very sparse (300 meter distance between shots), producing sampling artifacts which were attenuated (thought not completely) by the Hessian inversion. To have certainty that the artifact attenuation is the result of Hessian inversion and not from other process, regularization filters were turned off for these tests. Sometimes other approaches may work as well with less cost (directional filters, interpolation). We will not discuss those methods here, since we are interested in the benefits and problems for linearized inversion.

Adaptive data simplification

Now let us revisit the problem of modeling operators that do not accurately match the data. More importantly, let us discuss a possible solution that does not involve much more computing power than is normally available. To invert or eliminate the Hessian during inversion, LSMIG relies on the residual errors during iterations. Iterative algorithms, like conjugate gradient and others, use the residuals to calculate the direction and amplitudes (gradient and step size), of the necessary corrections to the reflectivity function or model. Failure of conditions a), b) or c) mentioned before leads the inversion to calculate incorrect model updates, which translate to noise in the image.

Clearly, we should try to fulfill these conditions by creating operators that can predict all our data correctly. However, in practice, these attempts can only succeed with very expensive prediction operators, e.g. visco-elastic. In particular, for industrial data sets of hundreds of gigabytes or terabytes, accurate modeling becomes unfeasible when it has to be applied many times in iterative methods. Alternatively, several simplifying approaches are commonly used:

1. Applying inversion model constraints to eliminate undesired features from the solution. This works well to compensate for failure of the condition b) above.
2. Pre-process the data, or data simplification, which is removing from the data events that the operator cannot predict. This is a very efficient manner to deal with failure of the condition c) above, but it requires us to know *a-priori* what these unpredictable elements are. That is why LSMIG industry successes usually involve very experienced geophysicists.

Now, we consider ways to apply *adaptive* data simplification. This approach is an alternative to straightforward data simplification that can use the evolution of the residuals as a tool to detect and turn off those elements the operator cannot predict.

The task of inverting the Hessian for surface seismic data is computationally expensive and because of that is usually done by iterative algorithms. In each iteration, the part of the data not properly predicted by the current reflectivity model is mapped back to the model space through the adjoint operator to calculate the reflectivity corrections required to improve fitting. The data residuals are calculated by taking the difference between data and predictions;

$$\mathbf{R} = \mathbf{d} - \mathbf{L}\mathbf{m}. \tag{10}$$

As all linearized inversions do, LSMIG assumes that $\mathbf{R}$ is completely produced by errors in $\mathbf{m}$. In reality, however, $\mathbf{R}$ has two components, one due to errors in $\mathbf{m}$ and another produced by errors in $\mathbf{L}$:

$$\mathbf{R} = \mathbf{L}(\boldsymbol{\Delta}\mathbf{m}) + (\boldsymbol{\Delta}\mathbf{L})\mathbf{m} = \mathbf{R_1} + \mathbf{R_2}. \tag{11}$$

When calculating reflectivity updates, we are interested in $\mathbf{R_1}$, but we only know $\mathbf{R}$. We need to find ways to minimize $\mathbf{R_2}$, or separate it from $\mathbf{R}$. In the next section we will discuss where $\mathbf{R_2}$ comes from.

To avoid introducing noise during inversion by mapping $\mathbf{R_2}$ from Eq. (11) into the model, the only correct solution is to improve the modeling operator. However, each attempt we make to match nature closely leads to orders of magnitude increases in computation cost. Here consider a compromise solution consisting of detecting R_2 in the residuals and removing it. It is common to monitor residual energy (one number) as a measure of convergence for the inversion. How this number changes with iterations tells us how well inversion is evolving. In addition, we can follow residual evolution with iterations as a full data space. As iterations proceed, we can detect residual components that are consistently increasing and turn them off. Further, we could use some prediction method (linear or non-linear) to detect data components that can't be properly matched by the operator. By masking from the inversion these residual components, the noise introduced by operator error in R_2 can be attenuated from the inversion.

To illustrate adaptive data simplification, we try a simple approach where we keep track of how individual residual points in data space change with iterations. If a particular residual point consistently increases (for example after two iterations), we multiply this residual point by a low data weight. A better method would be to follow events rather than points, what we haven't yet implemented that. Figure 13a shows an Kirchhoff depth migration for the Marmousi model. Figure 13b shows LS Kirchhoff after a few iterations. Complex events not predicted by Kirchhoff modeling because of differences between ray tracing and finite-difference modeling mapped into incorrect model updates and therefore noise in the image. Ray trace approximations in the deep central part of the Marmousi model produce many elements that can't be predicted properly. By detecting them and turning them off with a mask during the inversion (Figure 13c), we can stabilize the inversion on the poorly predicted part and achieve benefits where data can be predicted correctly (Figure 13d). Notice that there is some similarity between this approach and robust data inversion, but here we are using the evolution of residuals with iterations instead of just their energy with respect to the background as in outlier detection.

CONCLUSIONS

LSMIG, like other linearized inversions, requires an error-free operator to connect data and model spaces. Although desirable, in practice this is impossible to achieve. Even when working with synthetic finite difference data, LSMIG requires the use of Born or Kirchhoff modeling that can only partially predict the data. Even further, simple velocity errors or smoothness is sufficient to limit the quality of our predictions compared with data produced by Mother Nature. We have shown that we can apply model constraints and data simplification to alleviate this problem and improve LSMIG results. Although efficient, these approaches are often difficult to use because they require *a priori* knowledge of the mismatches between data and model. We have discussed some ways in which this can be achieve adaptively. Our approach is just one of many possibilities, probably not the most optimal, but we believe it serves as a proof of concept for inspiring further work on the topic.

ACKNOWLEDGMENTS

Our gratitude for CREWES sponsors for contributing to this seismic research. I also gratefully acknowledge support from NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13.

REFERENCES

Claerbout, J., 1992, Earth sounding analysis, Processing versus inversion: Blackwell Scientific Publications, Inc.

Hanke, M., and Hansen, P. C., 1993, Regularization methods for large-scale problems: Surv. Math. Ind, **3**, No. 4, 253–315.

Kuehl, H., 2002, Least-squares wave-equation migration/inversion:

Nemeth, T., Wu, C., and Schuster, G. T., 1999, Least squares migration of incomplete reflection data: GEOPHYSICS, **64**, No. 1, 208–221.

Schuster, G., 2017, Seismic Inversion: Society of Exploration Geophysicists.

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

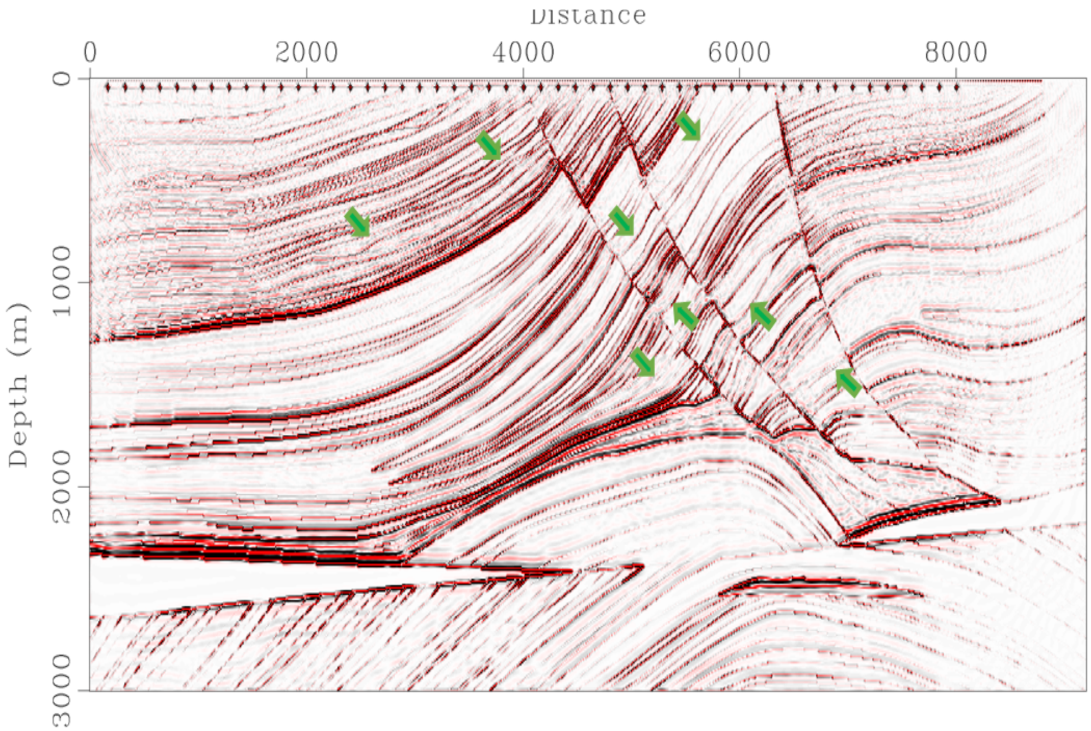

FIG. 5. RTM for Marmousi model

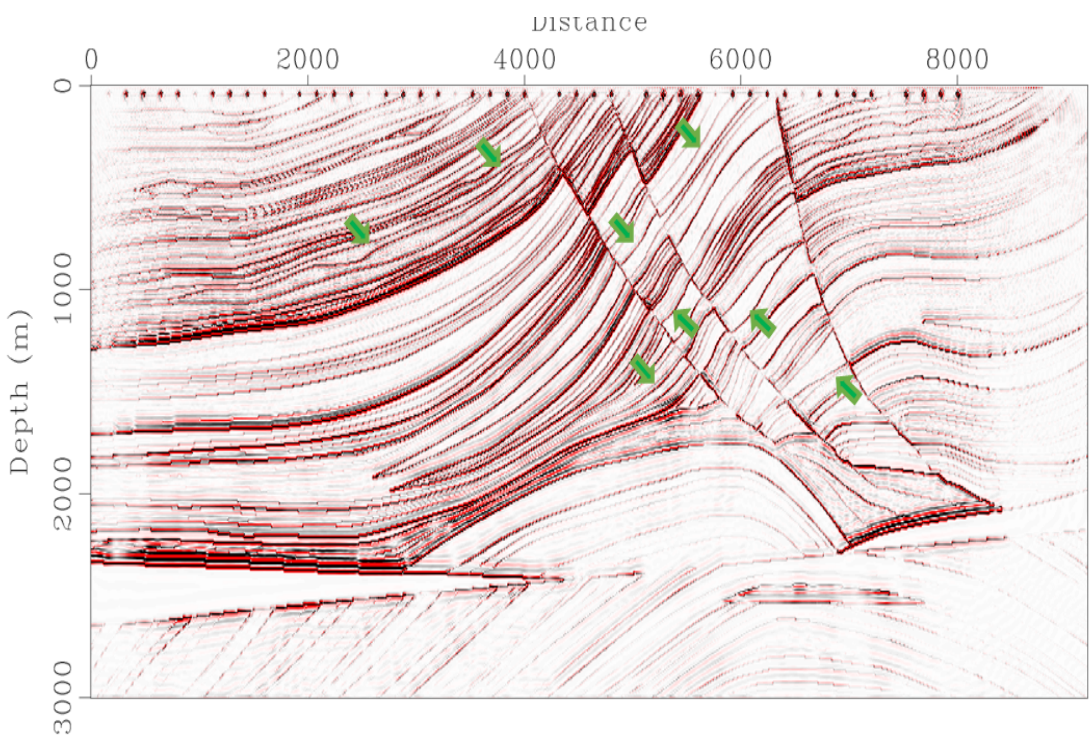

FIG. 6. LSRTM for Marmousi model. Arrows point to several improvements on focusing

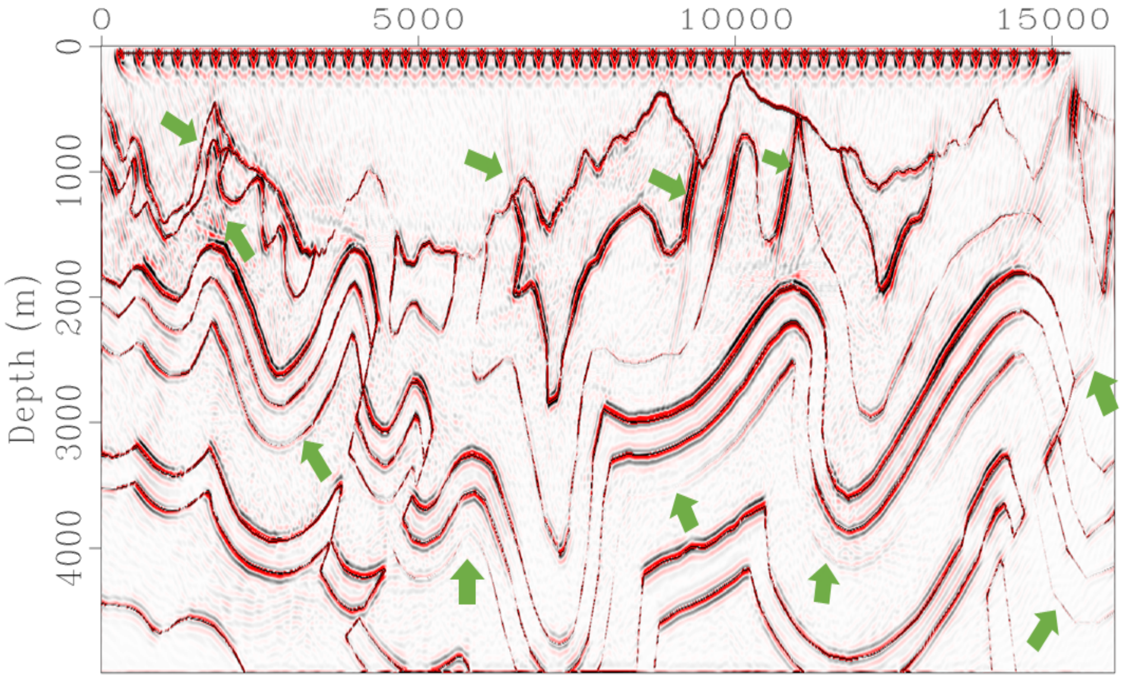

FIG. 7. RTM for Foothills model (model94 from Amoco)

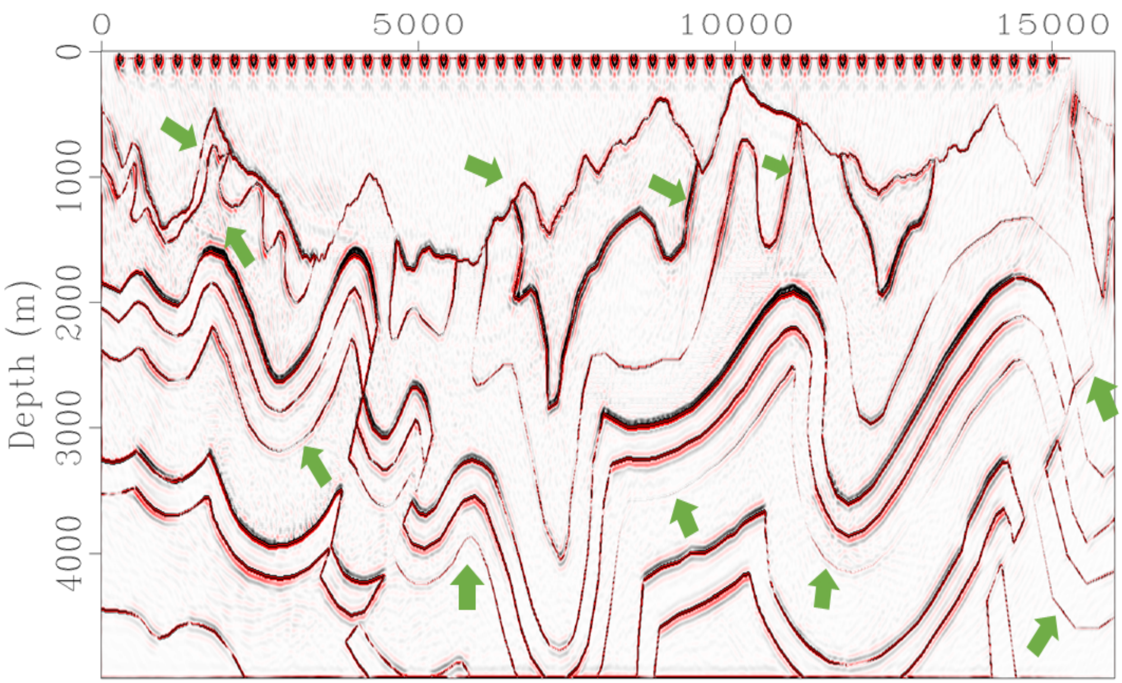

FIG. 8. LSRTM for model94. Arrows point to several improvements on focusing

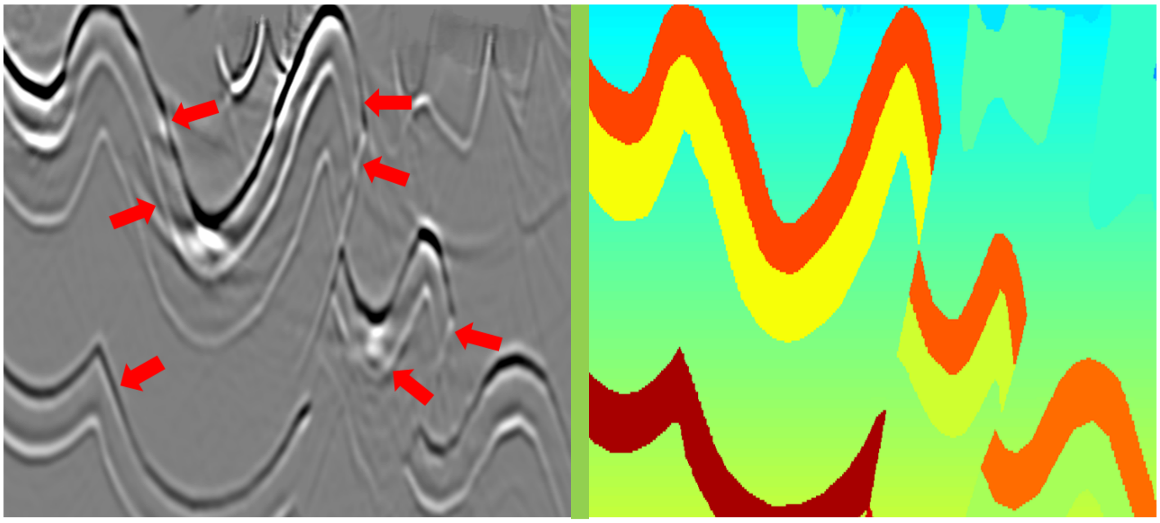

FIG. 9. Kirchhoff depth migration for model94 (left), velocity model (right)

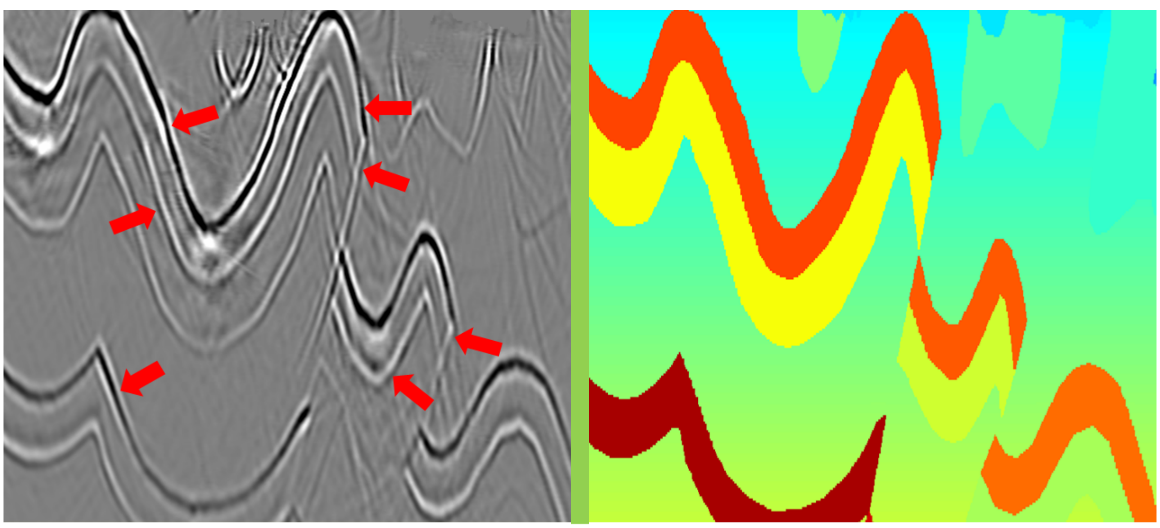

FIG. 10. LS Kirchhoff migration for model94 (left), velocity model (right). Arrows point to several improvements on focusing

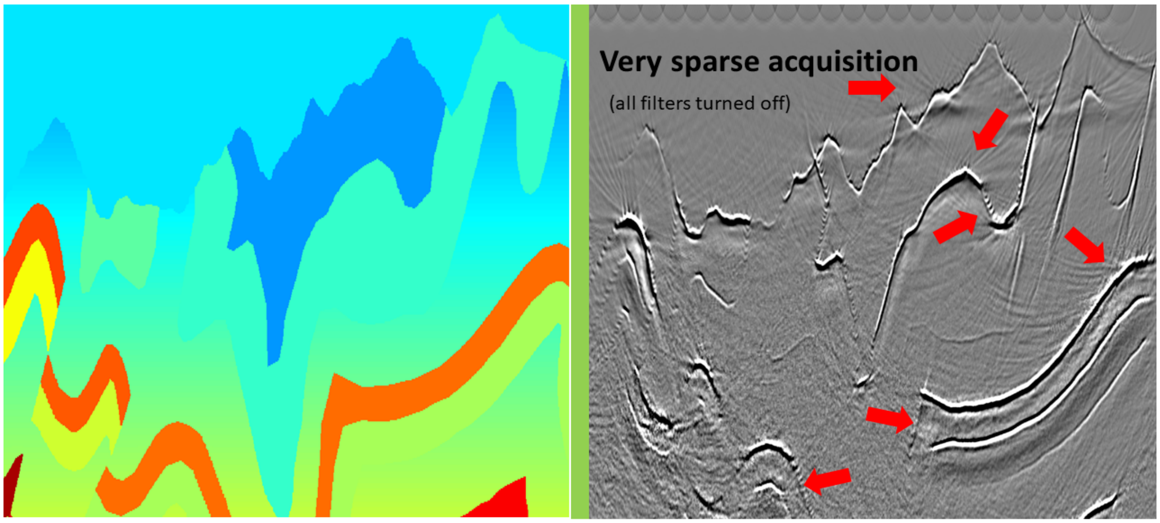

FIG. 11. Velocity model (left), Kirchhoff depth migration for model94 (right). Poor sampling due to a simulated very sparse survey produces many migration artifacts.

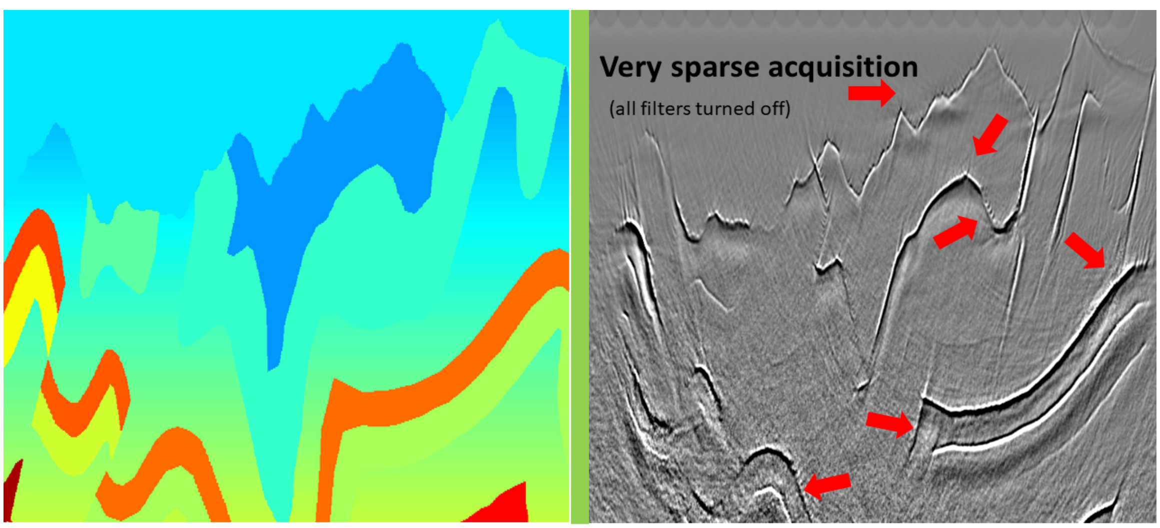

FIG. 12. Velocity model (left), LS Kirchhoff migration for model94 (left). Arrows point to migration artifacts being attenuated by the Hessian inversion (no filtering in this example)

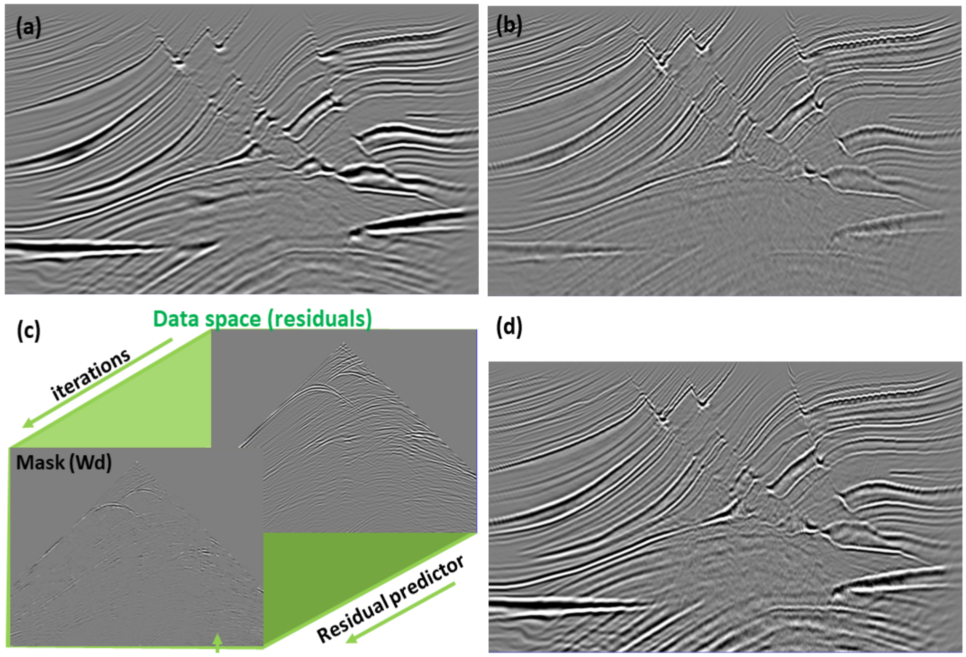

FIG. 13. a) Kirchhoff depth migration for Marmousi model, b) Kirchhoff LSMIG with noise introduced during iterations, c) Events in residuals that consistently increase during iterations are masked out, c) Kirchhoff LSMIG with adaptive data simplification.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation, vol. 89: siam.

Wang, J., Kuehl, H., and Sacchi, M. D., 2003, Least-squares wave-equation avp imaging of 3d common azimuth data, *in* SEG Technical Program Expanded Abstracts 2003, Society of Exploration Geophysicists, 1039–1042.