Least-squares reverse time migration with factorization-free priorconditioning

Gaurav Dutta[1], Matteo Giboli[2], Paul Williamson[2], and Gerard T. Schuster[1]
[1]*King Abdullah University of Science and Technology,* [2]*TOTAL Exploration and Production, France*

SUMMARY

We present a least-squares reverse time migration (LSRTM) method using a factorization-free priorconditioning approach to overcome the low signal-to-noise (SNR) problem arising out of using severely undersampled data. Priorconditioning is a technique where the prior information is incorporated directly into the forward operator and into the solution space of the problem. The prior information that is used in this work is that the inverted reflectivity is sparse in the radon domain. The proposed method is factorization-free since the forward mapping is defined through the action of a sparse operator on a vector. The priorconditioning method is shown to produce reliable images with good SNR and free from aliasing artifacts when using very sparse shots for both synthetic and field data.

INTRODUCTION

Least-squares migration (LSM) has been shown to produce images with balanced amplitudes, better resolution and fewer artifacts than standard migration (Lailly, 1984; Nemeth et al., 1999; Duquet et al., 2000; Plessix and Mulder, 2004; Dai and Schuster, 2009; Tang, 2009; Wong et al., 2011). Besides a reverse time migration (RTM) of the data residual, every iteration of LSRTM involves Born modeling to compute the predicted data from the reflectivity image and to estimate the steplength. Thus, every iteration of LSRTM is approximately 2-3 times more computationally expensive than that of standard RTM and the cost becomes proportional to the number of iterations carried out. Thus, the computational cost of LSRTM becomes very high for practical 3D problems.

To reduce the cost of standard RTM, Morton and Ober (1998) and Romero et al. (2000) proposed the idea of multisource phase-encoded migration where they blended several shotgathers using encoding functions with random time shifts and random source polarities into one supergather which they then migrated. Later, Dai et al. (2010) extended this idea to multisource LSRTM and showed that by an iterative migration of supergathers, multisource LSRTM can produce more accurate reflectivity images than standard RTM at a fraction of the computational cost. Herrmann and Li (2012) adopted a similar approach and used a combination of randomized dimensionality-reduction and divide-and-conquer-techniques to decimate the LSM problem as a series of smaller sub-problems where each sub-problem involved iterating on a small randomized subset of the data. They also combined their approach with compressive sensing and curvelet-domain sparse recovery (Candes et al., 2006) to obtain crosstalk free images from multisource LSM.

A simpler way of reducing the computational cost of LSRTM is by migrating very sparse shots since the cost of LSRTM is also proportional to the number of shotgathers migrated. However, using very sparse shots during migration has its pitfalls because the final image will be degraded in quality because of low SNR and migration artifacts which are not cancelled out by insufficient stacking. Thus, for incomplete or undersampled data, it becomes important to incorporate some sort of regularization into the inversion that would allow for a more accurate representation of the subsurface model parameters. Relaxing the sampling requirements will also lead to a reduction in the cost of data acquisition and processing.

If a regularization term is included in the misfit functional, the solutions exhibit very slow convergence. The regularization is controlled through a damping parameter which can be estimated from the L-shaped curve constructed by a log-log plot of the length of the model vector vs the length of the residual vector for different choices of the damping parameter (Calvetti et al., 2000). Such an approach becomes impractical for computationally expensive problems like full waveform inversion or LSM. A change of basis for the reflectivity using some linear transform or some form of model reparameterization (Harlan, 1995; Fomel and Guitton, 2006) are possible ways to account for the effect of the regularization term in the inversion without explicitly computing the damping parameter at every iteration. As mentioned by Kingsbury (2001), the choice of a suitable basis is dependent on 1) perfect reconstruction of the parameters after forward and inverse transforms, 2) suitability for use with conjugate gradient, 3) efficient computation, and 4) minimal redundancy. In our work, we use a high resolution local Radon transform of the reflectivity as a basis function and use a regularization term that imposes sparseness constraints on the inverted reflectivity in the Radon domain. Using a change of basis approach, we show that the prior or the regularization term gets embodied into a transformed forward operator and the prior information of sparse reflectivity is exploited without using an explicit regularization term. The benefits of priorconditioning is that the convergence is faster than the counterpart regularized problem and more meaningful updates can be seen in very few iterations.

THEORY

For a regularized L2-norm inverse problem, we seek to find a solution that best explains the observed data and is consistent with the prior knowledge that is available before any observations are made. In a Bayesian framework, this can be represented by the likelihood and the prior probability distributions, $P_{\mathbf{D}}(\mathbf{d}|\mathbf{m})$ and $P_{\mathbf{M}}(\mathbf{m})$, respectively. In the context of a seismic imaging problem, the subscripts $\mathbf{D}$ and $\mathbf{M}$ represent the data-space and the image-space, respectively, $\mathbf{d}$ represents the observed data and $\mathbf{m}$ represents the reflectivity model.

For a model with zero-mean Gaussian noise, these probability

DOI http://dx.doi.org/10.1190/segam2015-5912138.1

distributions can be expressed as

$$P_{\mathbf{D}}(\mathbf{d}|\mathbf{m}) \propto \exp\left\{-\frac{1}{2}||\mathbf{d}-\mathbf{Lm}||_2^2\right\},$$
$$P_{\mathbf{M}}(\mathbf{m}) \propto \exp\left\{-\frac{\lambda}{2}\mathbb{R}(\mathbf{m})\right\}, \qquad (1)$$

where $\mathbb{R}(\mathbf{m})$is a discrete regularizer/prior that imposes constraints on the solution $\mathbf{m}$. These constraints can be $\mathbf{m}$ should be sparse or the edges/reflectors in $\mathbf{m}$ should be sharp. Here, $\mathbf{L}$ represents a linear modeling operator and $\lambda > 0$ controls the strength of the regularization term. The maximum a-posteriori (MAP) estimate of the model can be expressed using Bayes' theorem as

$$\mathbf{m}_{MAP} = \underset{\mathbf{m}\in\mathbf{M}}{\arg\max}\ P(\mathbf{m}|\mathbf{d}),$$
$$= \underset{\mathbf{m}\in\mathbf{M}}{\arg\max}\ P_{\mathbf{D}}(\mathbf{d}|\mathbf{m})P_{\mathbf{M}}(\mathbf{m})\ , \qquad (2)$$

which can also be obtained by minimizing the misfit function given by

$$\underset{\mathbf{m}\in\mathbf{M}}{min}\left[\phi(\mathbf{m}) = \frac{1}{2}||\mathbf{d}-\mathbf{Lm}||_2^2 + \frac{\lambda}{2}\mathbb{R}(\mathbf{m})\right]. \qquad (3)$$

If we assume that $\lambda \neq 0$ and the prior $\mathbb{R}(\mathbf{m}) = \|\mathfrak{R}\mathbf{m}\|_1 = \|\mathbf{W}\mathfrak{R}\mathbf{m}\|_2^2 = ||\mathbf{Rm}||_2^2$ is some weighted transform that would make the image sparse in that domain, then the misfit function in equation 3 can be written as

$$\underset{\mathbf{m}\in\mathbf{M}}{min}\left[\phi(\mathbf{m}) = \frac{1}{2}||\mathbf{d}-\mathbf{Lm}||_2^2 + \frac{\lambda}{2}||\mathbf{Rm}||_2^2\right]. \qquad (4)$$

Here $\mathfrak{R}\mathbf{m}$ can be any transform operation that imposes sparseness constraints on the solution $\mathbf{m}$ and $\mathbf{W}$ is any model dependent weighting matrix. In this paper, we have taken $\mathfrak{R}\mathbf{m}$ to be the high-resolution Radon transform of the image $\mathbf{m}$ and $\mathbf{Rm}$ is the weighted-Radon transform of $\mathbf{m}$. The weights are model dependent and can be chosen to be varying locally or defined through a moving window over the samples in the model or can be estimated using quantile measures. In our present implementation, the weights are chosen locally.

The normal equations corresponding to the misfit function in equation 4 are given by

$$\left(\mathbf{L}^T\mathbf{L} + \lambda\mathbf{R}^T\mathbf{R}\right)\mathbf{m} = \mathbf{L}^T\mathbf{d}, \qquad (5)$$

which can be solved using any conjugate gradient based least-squares algorithm (LSQR type approach) (Paige and Saunders, 1982; Arridge et al., 2014). The functional $\phi(\mathbf{m})$ in equation 4 is minimized over the Krylov space

$$\kappa^{\mathbf{L}^T\mathbf{L}+\lambda\mathbf{R}^T\mathbf{R}} = \mathrm{span}\{\mathbf{L}^T\mathbf{d}, (\mathbf{L}^T\mathbf{L}+\lambda\mathbf{R}^T\mathbf{R})\mathbf{L}^T\mathbf{d}, \ldots,$$
$$(\mathbf{L}^T\mathbf{L}+\lambda\mathbf{R}^T\mathbf{R})^{i_{\max}-1}\mathbf{L}^T\mathbf{d}\} \subset \mathbf{M}, \qquad (6)$$

where $i_{\max}$ is the limit of the Krylov subspace in equation 6. As noted by Arridge et al. (2014), the difficulties associated with the solutions in the subspace described by equation 6 are 1) the solutions exhibit slow convergence, and 2) the inversion is controlled by the parameters λ and $i_{\max}$. If the damping parameter λ is estimated empirically, any change in λ will require a re-computation of the subspace defined by equation 6. If λ is evaluated from the L-shaped curve constructed from a log-log plot of the length of the model vector and the residual vector (Calvetti et al., 2000), it becomes impractical for a computationally expensive problem like full waveform inversion or least-squares migration.

Thus, to incorporate the information contained in the prior into the solution, a change of basis is required for the reflectivity $\mathbf{m}$. If $\mathbf{R}$ is invertible in equation 4, a change of basis

$$\hat{\mathbf{m}} = \mathbf{Rm},\ \hat{\mathbf{L}} = \mathbf{LR}^{-1}, \qquad (7)$$

in equation 4 gives

$$\underset{\hat{\mathbf{m}}\in\mathbf{RM}}{min}\left[\hat{\phi}(\hat{\mathbf{m}}) = \frac{1}{2}||\mathbf{d}-\hat{\mathbf{L}}\hat{\mathbf{m}}||_2^2 + \frac{\lambda}{2}||\hat{\mathbf{m}}||_2^2\right]. \qquad (8)$$

The normal equations corresponding to the misfit function in equation 8 are given by

$$\left(\hat{\mathbf{L}}^T\hat{\mathbf{L}} + \lambda\mathbf{I}\right)\hat{\mathbf{m}} = \hat{\mathbf{L}}^T\mathbf{d},$$
$$\left(\mathbf{R}^{-T}\mathbf{L}^T\mathbf{LR}^{-1} + \lambda\mathbf{I}\right)\hat{\mathbf{m}} = \mathbf{R}^{-T}\mathbf{L}^T\mathbf{d}. \qquad (9)$$

Since the prior information is now contained in the transformed operators, $\hat{\mathbf{L}}$ and $\hat{\mathbf{L}}^T$, we can set $\lambda = 0$ in equations 8 and 9 to get the normal equations,

$$\hat{\mathbf{L}}^T\hat{\mathbf{L}}\hat{\mathbf{m}} \approx \hat{\mathbf{L}}^T\mathbf{d},$$
$$\mathbf{R}^{-T}\mathbf{L}^T\mathbf{LR}^{-1}\hat{\mathbf{m}} \approx \mathbf{R}^{-T}\mathbf{L}^T\mathbf{d}, \qquad (10)$$

and the misfit function,

$$\underset{\hat{\mathbf{m}}\in\mathbf{RM}}{min}\left[\hat{\phi}(\hat{\mathbf{m}}) = \frac{1}{2}||\mathbf{d}-\hat{\mathbf{L}}\hat{\mathbf{m}}||_2^2\right]. \qquad (11)$$

The new functional $\hat{\phi}(\hat{\mathbf{m}})$ in equation 11 gets minimized over a different Krylov space given by

$$\kappa^{(\mathbf{R}^T\mathbf{R})^{-1}\mathbf{L}^T\mathbf{L}} = \mathrm{span}\{(\mathbf{R}^T\mathbf{R})^{-1}\mathbf{L}^T\mathbf{d}, ((\mathbf{R}^T\mathbf{R})^{-1}\mathbf{L}^T\mathbf{L})(\mathbf{R}^T\mathbf{R})^{-1}\mathbf{L}^T\mathbf{d},$$
$$\ldots((\mathbf{R}^T\mathbf{R})^{-1}\mathbf{L}^T\mathbf{L})^{i_{\max}-1}(\mathbf{R}^T\mathbf{R})^{-1}\mathbf{L}^T\mathbf{d}\} \subset \mathbf{RM}. \qquad (12)$$

On comparing the Krylov spaces in equations 6 and 12, it can be seen that the information contained in the prior/regularizer is now contained in the transformed operator $\hat{\mathbf{L}}$ in equations 10 and 11. Thus, the prior information can be exploited without using explicit regularization, i.e., by setting $\lambda = 0$ and letting the number of internal iterations in the conjugate gradient algorithm play the role of the regularizer. Even though the system of equations is only partially solved at every external iteration because the conjugate gradient algorithm is stopped before the solution is complete, the total matrix operator resulting from the priorconditioning is diagonalized enough so that one can avoid using a damping factor for the smaller eigenvalues. On the contrary, in equation 6, the prior information is controlled through the damping parameter λ which is difficult to estimate. Also, for different values of λ, the solutions in equation 12 do not change whereas the solutions in equation 6 need to be recomputed from the beginning.

DOI http://dx.doi.org/10.1190/segam2015-5912138.1

Such a priorconditioning approach is useful in solving a regularized L2 norm problem because it provides a way of incorporating the prior information directly into the forward operator and into the solution space of the problem. The Krylov space solutions in equation 12 clearly reveal that the inverted images from the very first iteration will have features with high prior probability. Building these features using an explicit regularization term will require a large number of iterations because the eigenvalues associated with them will be very small.

NUMERICAL RESULTS

Priorconditioned LSRTM is tested on the 3D Sandia/SEG C3 45 shot subset data. There are only 45 shots spread on a 5×9 source grid on the surface with a 960 m shot and shot-line separation. Each shot is recorded by a 201×201 receiver grid with a 50 m spacing between the receivers. Figure 1 compares the standard RTM, LSRTM and the priorconditioned LSRTM images after 10 iterations. The RTM image in Figure 1(a) suffers from very strong backscattering noise because of the presence of the salt body. The depth slices show strong acquisition footprint signatures and the reflector amplitudes are also very weak. The image contains significant high-frequency noise because of using severely undersampled data. The LSRTM image in Figure 1(b) shows some improvements over the standard RTM image. The reflector amplitudes are better balanced and the acquisition footprints in the depth slices are mitigated. However, the aliasing noise is still prominent and is severe below the salt body. The priorconditioned LSRTM image, shown in Figure 1(c), is free from the aliasing noise and the subsalt images are cleaner when compared to the standard RTM and LSRTM images. The salt boundaries are also better delineated in the crossline sections. The priorconditioned images also have a much better SNR than the standard RTM and LSRTM images.

The effectiveness of priorconditioned LSRTM is also demonstrated on a real OBS dataset. There are 26 OBS nodes at a spacing of 402.5 m. There are 360 shots fired at a spacing of 50 m on the surface. The common shot gathers are then sorted into common receiver gathers using reciprocity. Thus, for migration there are only 26 very sparse shots at a spacing of 402.5 m and each shot is recorded by 360 receivers spread at an interval of 50 m. The sea bottom is too shallow in this case, so mirror migration will not differ much from standard migration. The migration velocity model, shown in Figure 2, is obtained by early-arrival full waveform inversion.

During migration, data upto 3-10 Hz are migrated since the velocity tomogram was obtained by inverting for frequencies upto 10 Hz. The standard RTM and LSRTM images are shown in Figures 3(a) and 3(b), respectively. It is evident from these images that the aliasing artifacts are severe in the shallow parts. The sparse shot sampling has also led to low SNR in the images. The low SNR and aliasing artifacts make the interpretation of these images very difficult. The priorconditioned RTM and LSRTM images, shown in Figures 3(c) and 3(d), have fewer aliasing artifacts and a much better SNR than the standard RTM and LSRTM images. The reflectors in the shallow parts can be clearly delineated and the reflector amplitudes are very well balanced.

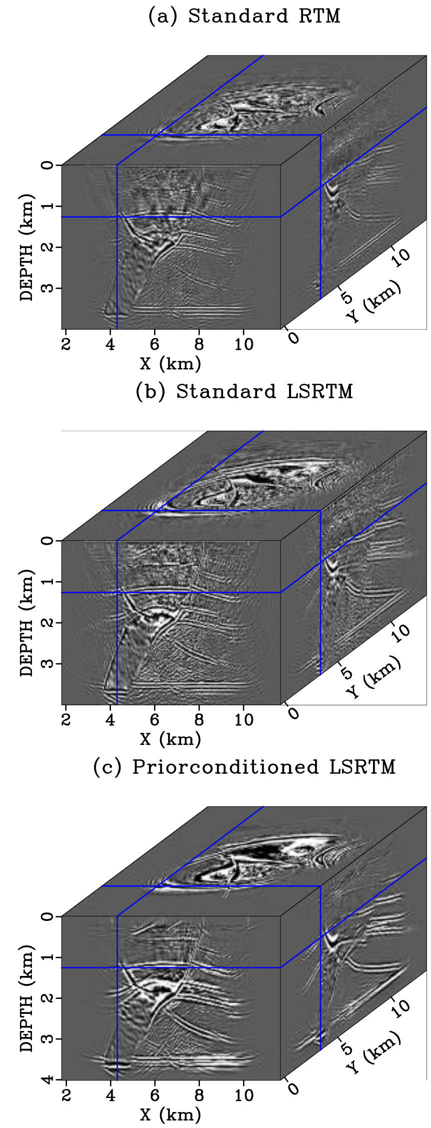

Figure 1: Comparison between images from (a) standard RTM, (b) standard LSRTM, and (c) priorconditioned LSRTM.

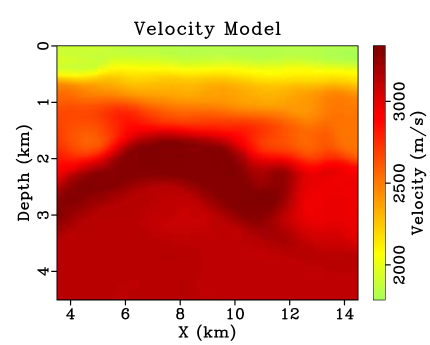

Figure 2: Migration velocity model for the OBS data obtained using Gauss Newton FWI.

DOI http://dx.doi.org/10.1190/segam2015-5912138.1

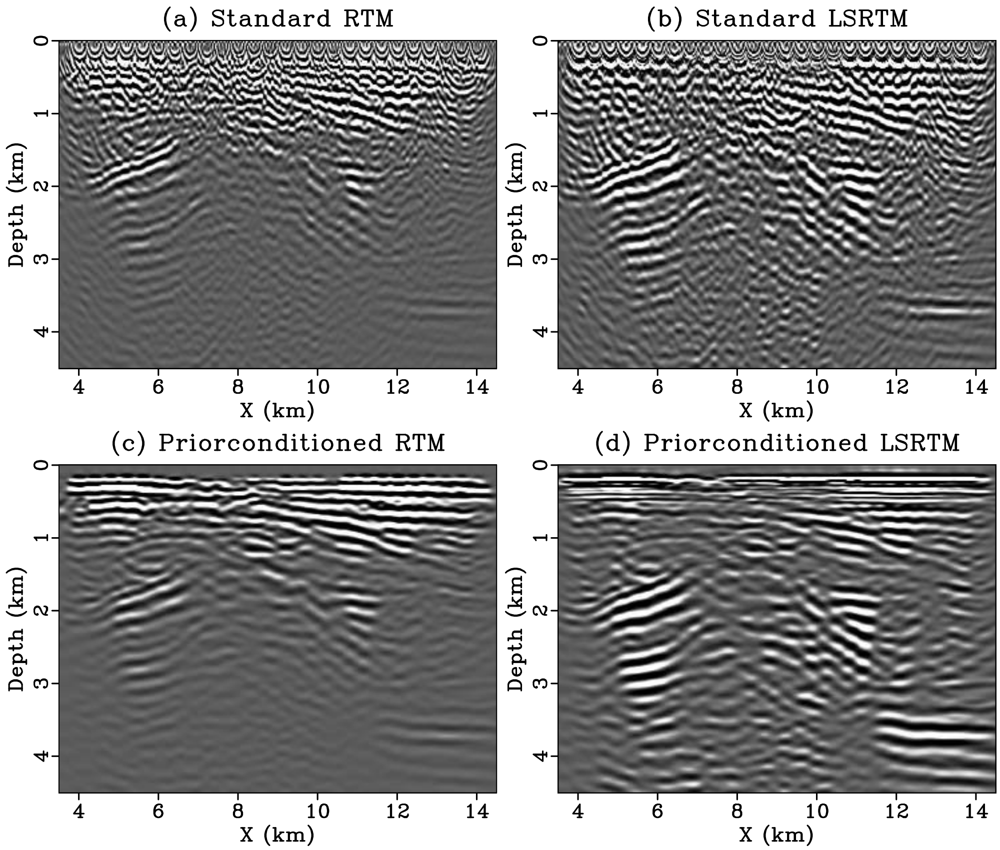

Figure 3: Comparison between images from (a) standard RTM, (b) standard LSRTM, (c) priorconditioned RTM, and (d) priorconditioned LSRTM. The LSRTM images are obtained after 10 iterations.

CONCLUSIONS

A priorconditioning approach is presented for LSRTM that incorporates a prior/regularizer directly into the forward operator and into the solution space of the problem. This approach is factorization-free since the forward mapping is defined through the action of a sparse operator on a vector. This is particularly useful for time-domain LSM problems where the forward and adjoint mapping cannot be explicitly computed and stored as a matrix. Priorconditioning also requires a fewer number of iterations to converge when compared to using an explicit regularization term in the objective function. Our numerical tests on synthetic and field data validate that priorconditioning can be used to reduce the computational cost of LSRTM by using very sparse shots and overcome the aliasing artifacts caused because of using severely undersampled data. A disadvantage of our present implementation using weighted radon transform with local weights is that it is sensitive to the presence of free-surface multiples or internal multiples in the data. This is because the false reflections from these multiples will have a prominent dip signature which might get boosted up if the weights are not chosen appropriately. Investigating the selection of appropriate weights in the presence of multiples is a topic of ongoing research.

ACKNOWLEDGEMENTS

The first author would like to thank the sponsors of the Center for Subsurface Imaging and Fluid Modeling (CSIM) consortium for their support and Abdullah AlTheyab for providing the velocity tomogram for the OBS data. We are also grateful to TOTAL E&P France for permission to publish this work. The plots for this abstract have been prepared using the Madagascar open-source software package.

DOI http://dx.doi.org/10.1190/segam2015-5912138.1

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2015 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Arridge, S., M. Betcke, and L. Harhanen, 2014, Iterated preconditioned LSQR method for inverse problems on unstructured grids: Inverse Problems, **30**, no. 7, 075009. http://dx.doi.org/10.1088/0266-5611/30/7/075009.

Calvetti, D., S. Morigi, L. Reichel, and F. Sgallari, 2000, Tikhonov regularization and the l-curve for large discrete ill-posed problems: Journal of Computational and Applied Mathematics, **123**, no. 1–2, 423–446. http://dx.doi.org/10.1016/S0377-0427(00)00414-3.

Candès, E., L. Demanet, D. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: Multiscale Modeling & Simulation, **5**, no. 3, 861–899. http://dx.doi.org/10.1137/05064182X.

Dai, W., C. Boonyasiriwat, and G. T. Schuster, 2010, 3D multisource least-squares reverse-time migration: 80th Annual International Meeting, SEG, Expanded Abstracts, **29**, 3120–3124.

Dai, W., and G. T. Schuster, 2009, Least-squares migration of simultaneous sources data with a deblurring filter: Presented at the 79th Annual International Meeting, SEG.

Duquet, B., K. J. Marfurt, and J. A. Dellinger, 2000, Kirchhoff modeling, inversion for reflectivity, and subsurface illumination: Geophysics, **65**, 1195–1209. http://dx.doi.org/10.1190/1.1444812.

Fomel, S., and A. Guitton, 2006, Regularizing seismic inverse problems by model reparameterization using plane-wave construction: Geophysics, **71**, no. 5, A43–A47. http://dx.doi.org/10.1190/1.2335609.

Harlan, W. S., 1995, Regularization by model reparameterization: Citeseer.

Herrmann, F. J., and X. Li, 2012, Efficient least-squares imaging with sparsity promotion and compressive sensing: Geophysical Prospecting, **60**, no. 4, 696–712. http://dx.doi.org/10.1111/j.1365-2478.2011.01041.x.

Kingsbury, N., 2001, Complex wavelets for shift invariant analysis and filtering of signals: Applied and Computational Harmonic Analysis, **10**, no. 3, 234–253. http://dx.doi.org/10.1006/acha.2000.0343.

Lailly, P., 1984, Migration methods: Partial but efficient solutions to the seismic inverse problem, *in* A. Santosa, Y.-H. Pao, and W. W. Symes, Inverse problems of acoustic and elastic waves: SIAM, 51, 1387–1403.

Morton, S. A., and C. C. Ober, 1998, Faster shot record depth migrations using phase encoding: 68th Annual International Meeting, SEG, Expanded Abstracts, 298, 1131–1134.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221. http://dx.doi.org/10.1190/1.1444517.

Paige, C. C., and M. A. Saunders, 1982, LSQR: An algorithm for sparse linear equations and sparse least squares: ACM Transactions on Mathematical Software, **8**, no. 1, 43–71. http://dx.doi.org/10.1145/355984.355989.

Plessix, R.-E., and W. A. Mulder, 2004, Frequency-domain finite-difference amplitude-preserving migration: Geophysical Journal International, **157**, no. 3, 975–987. http://dx.doi.org/10.1111/j.1365-246X.2004.02282.x.

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436. http://dx.doi.org/10.1190/1.1444737.

Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded Hessian: Geophysics, **74**, no. 6, WCA95–WCA107. http://dx.doi.org/10.1190/1.3204768.

Wong, M., S. Ronen, and B. Biondi, 2011, Least-squares reverse time migration/inversion for ocean bottom data: A case study: 81st Annual International Meeting, SEG, Expanded Abstracts, 30, 2369–2373.