# A short note on rank-2 relaxation for waveform inversion

*Augustin Cosse*[1,*], *Stephen D. Shank*[2], *and Laurent Demanet*[2]
[1] *Harvard, SEAS and UCL, ICTEAM.*
[2] *Department of Mathematics and Earth Resources Laboratory, Massachusetts Institute of Technology.*

## SUMMARY

This note is a first attempt to perform waveform inversion by utilizing recent developments in semidefinite relaxations for polynomial equations to mitigate non-convexity. The approach consists in reformulating the inverse problem as a set of constraints on a low-rank moment matrix in a higher-dimensional space. While this idea has mostly been a theoretical curiosity so far, the novelty of this note is the suggestion that a modified adjoint-state method enables algorithmic scalability of the relaxed formulation to standard 2D community models in geophysical imaging. Numerical experiments show that the new formulation leads to a modest increase in the basin of attraction of least-squares waveform inversion.

## INTRODUCTION

Imaging in the acoustic regime is traditionally posed as the full-waveform inversion (FWI) problem (see, e.g., Tarantola and Valette (1982) and Tarantola et al. (1984))

$$\begin{aligned} &\min_{m} \quad \|Su-d\|^2, \\ &\text{s.t} \quad \Delta u + \omega^2 m(x) u = f \qquad on \quad \Omega. \end{aligned} \tag{1}$$

Here $u$ is the wavefield, $d$ is the data vector, $S$ denotes the operator sampling the field at the receivers, and $m = 1/c^2(x)$ is the model reflectivity (the density is assumed constant). Both the frequency $\omega$ and the source term $f$ range over a large number of instances. We will further assume that the reflectivity can be decomposed into a known contribution $\overline{m}$ and an unknown perturbation $\tilde{m}$ such that $m(x) = \overline{m}(x) + W\tilde{m}(x)$, where $W$ is a window zeroing out a neighbourhood around the receivers.

The problem is hard when the lack of low frequencies in the data exacerbate the non-convexity of the objective as an implicit function of $m$. For instance, Gauthier et al. (1986) documents the importance of a good starting model. To date, the most successful ideas to determine a satisfactory starting model involve model extension in a higher-dimensional space, and data redundancy for further discrimination. Differential semblance optimization is one such approach (Symes (1993); Shen (2004); Symes and Carazzone (1991)), and wave equation migration velocity analysis is another one (Sava and Biondi (2004a,b); Liu and Bleistein (1995)). The approach presented in this note can be seen as an "algebraic" instance of extended modeling. Other interesting approaches to extend the basin of attraction of waveform inversion can be found in Haber et al. (2000); van Leeuwen and Herrmann (2013); Wu et al. (2013).

## CONVEX RELAXATION

Once discretized using an appropriate finite difference scheme, problem (1) can be expressed as a set of polynomial equations in the variables $m_1,\dots,m_N$, $u_1,\dots u_L$, where $N = n^2$ denotes the size of the grid and $L = N \times$ sources $\times$ frequencies. For notational clarity only, we write the derivation for a single source and a single frequency. Let $K$ denote the discrete Laplacian with absorbing boundary layers. We introduce the following convex relaxation of the set of polynomial equations (see Lasserre (2001); Laurent (2009)):

$$\begin{aligned} \min \quad & \|SX_{31} - d\|^2 \\ \text{s.t} \quad & X_{11} = 1 \\ & KX_{31} + \omega^2(\overline{m} \circ X_{31} + W\,\mathrm{diag}(X_{32})) = f, \\ & X \succeq 0, \end{aligned} \tag{2}$$

where the matrix $X$ is defined blockwise as

$$X = \begin{pmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{pmatrix}, \tag{3}$$

and $X \succeq 0$ means that $X$ is positive semi-definite. The intention is for $X$ to be a proxy for the rank-1 matrix

$$X_0 = (1 \quad \tilde{m}^* \quad u^*)^*\,(1 \quad \tilde{m}^* \quad u^*),$$

which we want to recover; here $X^*$ denotes the conjugate transpose of a matrix $X$. For large-scale applications such as the ones usually studied in geophysics, formulation (2) is clearly intractable since it requires consideration of $O(N^2)$ variables. Nonetheless, it is possible to get a computationally feasible formulation by introducing a low-rank factorization $RR^*$ for the matrix $X$, as proposed by Burer and Monteiro (2003):

$$\begin{aligned} \min \quad & \|S(R_3R_1^*) - d\|^2 \\ \text{s.t} \quad & (R_1R_1^*) = 1 \\ & K(R_3R_1^*) + \omega^2(\overline{m} \circ (R_3R_1^*) + W\,\mathrm{diag}(R_3R_2^*)) = f, \end{aligned} \tag{4}$$

where $R = (R_1^* \quad R_2^* \quad R_3^*)^*$. We would then minimize the augmented Lagrangian of problem (4) through gradient descent iterations. The drawback of such a descent is that the Helmholtz equation would be solved implicitly through the iterations rather than explicitly, which results in slow convergence. Furthermore, the relaxation in problem (2) may be too loose, and as a result introduce spurious high-rank minimizers. In order to write a truly tractable relaxed program, we propose to further constrain (4) and solve it efficiently by the consideration of an appropriate adjoint field. Note that if the rank of $R$ is 1, then problem (4) reduces to FWI.

# A short note on rank-2 relaxation for waveform inversion

## THE ADJOINT-STATE METHOD

Early instances of the adjoint-state formalism are found in Lax et al. (1957), Lions and Magenes (1972), and Chavent and Lemonnier (1974). Extensions and applications to geophysics include Tarantola (1984, 2005). A review is in Plessix (2006).

The adjoint-state method is helpful in that it provides the expression of the gradient for FWI without having to compute or store the Fréchet derivatives of the field $\delta_m u$. Let $H := \Delta + \omega^2 m(x)$. The equations of the adjoint-state method for problem (1) follow from minimizing the Lagrangian

$$\mathscr{L}(u,\tilde{m},q) = Re\left\{\|Su-d\|^2 - \langle q, H(\tilde{m})u - f\rangle\right\},$$

by moving along the gradient direction in $\tilde{m}$, while setting the partials with respect to $q$ and $u$ to 0. These zero-derivative conditions respectively imply the Helmholtz equation $Hu = f$ for the (state) field $u$, and the adjoint equation $H^*q = w$ for the (adjoint state) field $q$, where $w = S^*(Su-d)$ is the residual. An imaging condition then links $u$ and $q$ to form the model update. See the review by Symes (2009) for more details on the history and the mathematics of the adjoint-state method.

### Rank-$r$ formulation

This section introduces a version of the adjoint-state method to deal with problem (4), when the matrix $R$ (hence $X = RR^*$) has fixed rank $r > 1$. From now on, we assume that $R \in \mathbb{C}^{h\times r}$, $h = 1+N+L$. Denote its blocks by $R_1 = [\alpha_1,\dots,\alpha_r] \in \mathbb{R}^{1\times r}$, $R_2 = [m_1,\dots,m_r] \in \mathbb{R}^{N\times r}$, $R_3 = [u_1,\dots,u_r] \in \mathbb{C}^{L\times r}$, which then leads to a representation of $X$ as a sum of rank-1 matrices,

$$X \approx RR^* = \sum_{\ell=1}^{r} \begin{pmatrix} \alpha_\ell^2 & \alpha_\ell\tilde{m}_\ell^* & \alpha_\ell u_\ell^* \\ \alpha_\ell\tilde{m}_\ell & \tilde{m}_\ell\tilde{m}_\ell^* & \tilde{m}_\ell u_\ell^* \\ \alpha_\ell u_\ell & u_\ell\tilde{m}_\ell^* & u_\ell u_\ell^* \end{pmatrix}.$$

We then apply the constraints of formulation (2) to this last expression. This leads to the following problem, revealing a rank-$r$ version of the Helmholtz equation,

$$\begin{aligned} \min \quad & \frac{1}{2}\|\sum_{\ell=1}^{r}\alpha_\ell Su_\ell - d\|^2 \text{ s.t. } \sum_{\ell=1}^{r}\alpha_\ell^2 = 1, \\ & \sum_{\ell=1}^{r}\alpha_\ell Ku_\ell + \omega^2\alpha_\ell\overline{m}\circ u_\ell + \omega^2 W\tilde{m}_\ell\circ u_\ell = f. \end{aligned} \tag{5}$$

### Least-squares and gradient computations

The adjoint-state framework cannot directly be applied to formulation (5) because the rank-$r$ Helmholtz equation is underdetermined, and therefore does not possess a unique solution.

Let $L_\ell(\alpha_\ell,\tilde{m}_\ell) := \alpha_\ell K + \omega^2(\alpha_\ell\overline{m} + W\tilde{m}_\ell)$, and $L = [L_1,\dots,L_r]$, so that the rank-$r$ Helmholtz equation reads $Lu = f$. Explicitly,

$$Lu = \sum_{\ell=1}^{r}\alpha_\ell Ku_\ell + \omega^2(\alpha_\ell\overline{m}\circ u_\ell + W\tilde{m}_\ell\circ u_\ell).$$

A natural simplification is to consider the underdetermined least-squares (LS) solution to $Lu = f$. This is accomplished by requiring $u_\ell = L_\ell^* v$, and finding $v$ by solving the normal equation $\tilde{H}v = f$, where $\tilde{H} := \sum_{\ell=1}^{r} L_\ell L_\ell^*$. The corresponding reformulation of problem (5) in terms of $v$ reads

$$\begin{aligned} \min \quad & J(\alpha,v) = \|\sum_{\ell=1}^{r}\alpha_\ell SL_\ell^* v - d\|^2 \\ \text{s.t} \quad & \tilde{H}v = f, \qquad \sum_{\ell=1}^{r}\alpha_\ell^2 = 1. \end{aligned} \tag{6}$$

Introducing the adjoint field $q = q_R + iq_I$, $i = \sqrt{-1}$, the Lagrangian and its partials are given by

$$\mathscr{L}(\alpha,v,\tilde{m};q) = J(\alpha,v) - Re\left\{\langle q,\ \tilde{H}(\alpha,\tilde{m})v - f\rangle\right\},$$

$$\frac{\partial\mathscr{L}}{\partial\alpha} = \frac{\partial J}{\partial\alpha} - Re\langle q,\ \frac{\partial\tilde{H}v}{\partial\alpha}\rangle, \tag{L1}$$

$$\frac{\partial\mathscr{L}}{\partial\tilde{m}} = \frac{\partial J}{\partial\tilde{m}} - Re\langle q,\ \frac{\partial}{\partial\tilde{m}}\tilde{H}v\rangle, \tag{L2}$$

$$\frac{\partial\mathscr{L}}{\partial Re\{v\}} = \frac{\partial J}{\partial Re\{v\}} - Re\{(\tilde{H}^*q)\}, \tag{L3}$$

$$\frac{\partial\mathscr{L}}{\partial Im\{v\}} = \frac{\partial J}{\partial Im\{v\}} - Im\{(\tilde{H}^*q)\}. \tag{L4}$$

Following the traditional approach, we set (L3) and (L4) to zero. Let $\tilde{\alpha}$ and $\tilde{S}$ denote the operators $\tilde{\alpha} = [\alpha_1 I_N,\dots\alpha_r I_N]$ and $\tilde{S} = I_r \otimes S$, respectively. We can derive the solution of the adjoint equation as

$$q = (\tilde{H}^*)^{-1}(\tilde{\alpha}\tilde{S}L^*)^*(\tilde{\alpha}\tilde{S}L^*v - d).$$

The partials of $\tilde{H}$ with respect to $\tilde{m}_\ell$ and $\alpha_\ell$, $1 \le \ell \le r$ can be expressed as

$$\begin{aligned} \langle q, \partial_{\alpha_\ell}\tilde{H}\rangle =& 2\alpha_\ell\langle q, K(K^*v)\rangle + \langle q, 2\omega^4 W\tilde{m}_\ell\circ\overline{m}\circ v\rangle \\ & + \alpha_\ell\langle q, 2\omega^4\overline{m}^2\circ v\rangle \\ & + \langle q, K(\omega^2 W\tilde{m}_\ell\circ v)\rangle + 2\alpha_\ell\langle q, K(\omega^2\overline{m}\circ v)\rangle \\ & + \langle q, \omega^2 W\tilde{m}_\ell\circ K^*v\rangle + 2\alpha_\ell\langle q, \omega^2\overline{m}\circ K^*v\rangle. \\ \langle q, \partial_{\tilde{m}_\ell}\tilde{H}\rangle =& \overline{q}\omega^4 2\alpha_\ell\circ\overline{m}\circ Wv + 2\overline{q}\omega^4\circ W^2\tilde{m}_\ell\circ v + \\ & \alpha_\ell\overline{(K^*(q))}\omega^2\circ Wv + \overline{q}\omega^2\circ W\alpha_\ell\Delta^* v. \end{aligned}$$

The partials of the objective $\partial_{\alpha_\ell}J$, $\partial_{\tilde{m}_\ell}J$ are given by $\partial_{\alpha_\ell}J = 2Re\{\langle SL_\ell^*v, \sum_\ell \alpha_\ell SL_\ell^*v - d\rangle\}$; and $\partial_{\tilde{m}_\ell}J = 0$.

## ALGORITHM

The rank-2 adjoint-state method (henceforth R2AS) is summarized in Algorithm 1. We restrict to rank $r = 2$ since no substantial improvement was observed at higher ranks. We use coordinate descent, first minimizing over all $\tilde{m}_\ell$, followed by minimizing over all $\alpha_\ell$. These minimizations are performed using LBFGS (see Liu and Nocedal (1989)) from the knowledge of the gradients (L1) and (L2). We then normalize the $\alpha_\ell$ to ensure that they reside on the unit sphere. Finally, the approximation for $m$ is extracted by computing the leading eigenvector of $RR^*$, appropriately scaled.

**Algorithm 1** Rank-2 Adjoint State (R2AS)

**Input:** Initial iterates $\alpha^{(0)} \in \mathbb{R}^2$, $\tilde{m}^{(0)} \in \mathbb{R}^{N\times 2}$
**Output:** Approximate reflectivity $\tilde{m}^\star$

1: **while** $\|\sum_{\ell=1}^{2} \alpha_\ell S L_\ell^* u_\ell - d\|/\|d\| > \varepsilon$ **do**
2: Compute the LS field : $v = \tilde{H}^{-1} f$
(Forward Step)
3: LBFGS step in $\tilde{m}$
4: LBFGS step in $\alpha$
(Backward step)
5: Projection of $\alpha$ onto the unit sphere:
$\alpha \leftarrow \alpha/\|\alpha\|$.
6: Compute $u$ as $u = L^* v$.
7: **end while**
8: Obtain $\tilde{m}^\star$ as
$RR^* = \lambda_1 v_1 v_1^* + \lambda_2 v_2 v_2^* \qquad \tilde{m}^\star = (\sqrt{\lambda_1} v_1)_2$

### Numerical Experiments

All numerical experiments are conducted by adding 0.1% additive Gaussian white noise to the data. The rank-2 method is initialized by taking as first column an initial guess $\tilde{m}^{(0)}$ for which the adjoint-state method may fail, and as second column an uninformative perturbation (arbitrarily chosen as an oscillatory bump $g$, times a scalar $\beta$) of the same vector, i.e., $R^{(0)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ \tilde{m}^{(0)} & \tilde{m}^{(0)} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \beta g \end{pmatrix}$.

### Example 1: Camembert

We first benchmark our method on a version of the Camembert example (see Gauthier et al. (1986) and Fig.1). The domain is $80 \times 80$ (grid points) with a background reflectivity of 1, contrast $\kappa = \max(1/c^2) - \min(1/c^2)$ ranging from .5 to .6, and a narrow frequency band between $1.5/(2\pi)$ and $1/\pi$. Receivers and sources are placed all around the domain.

### Example 2: Shepp-Logan phantom

Full-waveform inversion is initialized with background $\overline{m} = (1-W)m_0$, $\tilde{m}^{(0)} = 1$. The domain is of size $120 \times 120$ (grid points) and the reflectivity constrast varies between .8 and 1. The frequencies are taken equispaced between $1.5/(2\pi)$ and $1/\pi$. The receivers and sources are placed all around the domain as shown in Fig. 4.

### Example 3: Marmousi 1

We finally apply our method to the Marmousi model (see Versteeg (1994)) at frequencies ranging from about 3.3 to about 6.4 Hz. This is in accordance with usual simulations (see for example Sirgue and Pratt (2004)). Receivers and sources are positioned at the surface (see Fig. 2, top). The initial guess is obtained by smoothing the original image with a Gaussian kernel of varying widths. Traditional full-waveform inversion starts failing for smoothing above 35 grid points. Our method recovers the right map for a smoothing of up to 40 grid points. For the best results, $\alpha_1$ and $\alpha_2$ should be chosen to be very different.

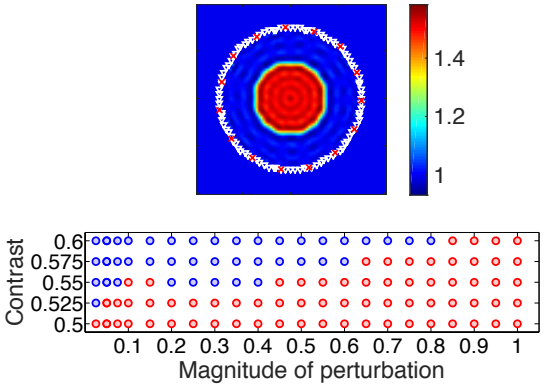


Figure 1: Top: Setup and recovery of R2AS for the Camembert example (reflectivity map with contrast $\kappa = .51$). Bottom: Performance of R2AS on the Camembert map for varying magnitudes of perturbation to the second column $\beta$ and contrasts $\kappa$. A red dot indicates convergence for the corresponding pair of parameters, while a blue dot indicates failure.

### Results

The R2AS approach leads to an increase of 5% to 10% in the basin of attraction of FWI for each model considered here. This improvement is measured with respect to the contrast for the Camembert and Shepp-Logan models, and with respect to the amount of smoothing of the Marmousi model.

The bottom portion of Fig. 1 plots the magnitude of the perturbation to the second column $\beta$ against the contrast of the Camembert model. Higher contrasts are more difficult to handle, and FWI gets trapped in spurious local minima when $\kappa \geq .51$. A higher $\beta$ intuitively represents a stronger push of the matrix $X$ away from the rank-1 manifold. Success at rank 2 and higher may be attributed to the ability of escaping the rank-1 local minima by operating in a higher-dimensional space.

## CONCLUSION

We have presented a variant of the adjoint-state method to efficiently solve the rank-2 moment relaxation of FWI. The improvement in local convexity around the global minimizer is intuitively explained by the ability to find better paths around local minima when performing the descent in a higher-dimensional space. The proposed approach yields a 5-10% enlargement of the basin of attraction for standard 2D community models.

## ACKNOWLEDGMENTS

AC thanks the Belgian National Science Foundation (FNRS) and MISTI Belgium for support. LD and SS thank Total SA. LD is also supported by AFOSR, ONR, and NSF. AC would like to thank Yunyue Elita Li and Leonardo Zepeda-Núñez for interesting discussions.

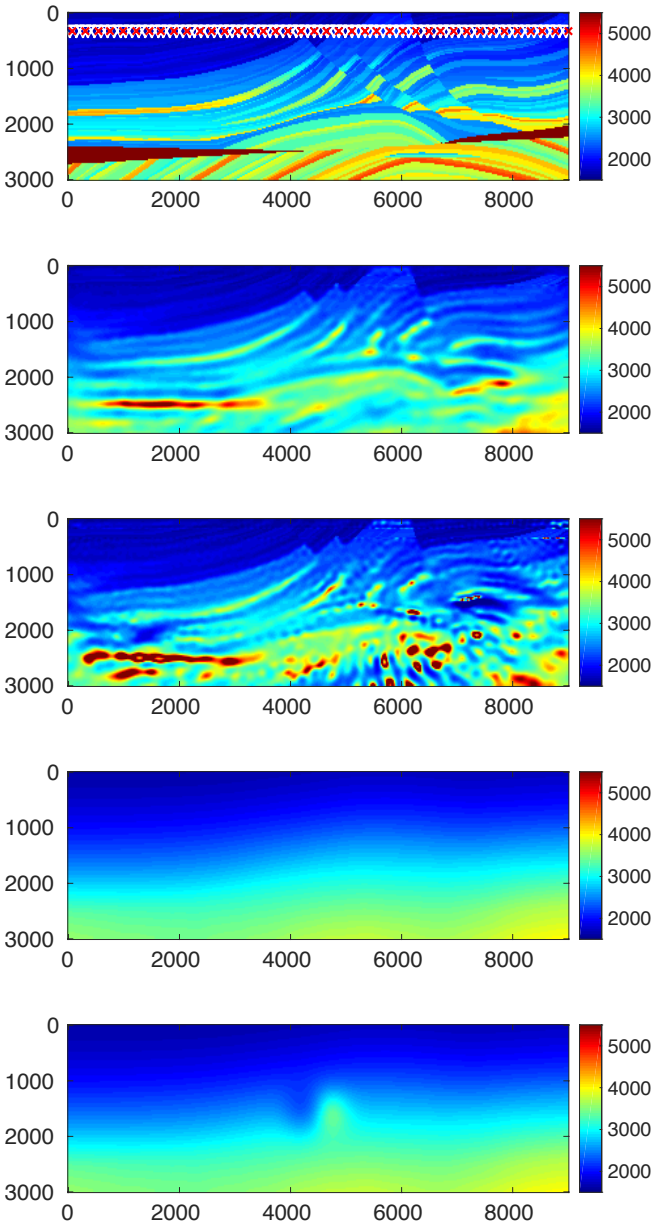


Figure 2: Marmousi. From top to bottom: Original configuration (sources and receivers are placed at the surface, and depicted by red crosses and white triangles, respectively); image recovered from rank-2 relaxation; image recovered through traditional least-squares inversion; initial $u$ component of the first column of $R$ (as well as initial guess for FWI); and initial $u$ component of the second column of $R$.

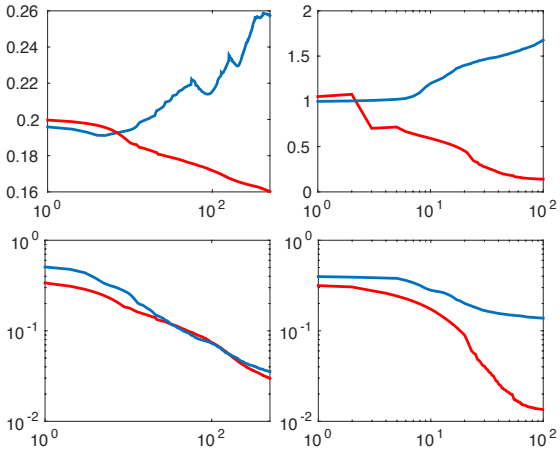


Figure 3: Top: Relative error for the Marmousi map (L) and the Shepp-Logan map (R). The error curve corresponding to R2AS is shown in red, the curve corresponding to imaging through full-waveform inversion is shown in blue. Bottom: Misfit for the Marmousi map (L) and Shepp-Logan map (R). Again the curves corresponding to R2AS are shown in red, the ones corresponding to full-waveform inversion in blue.

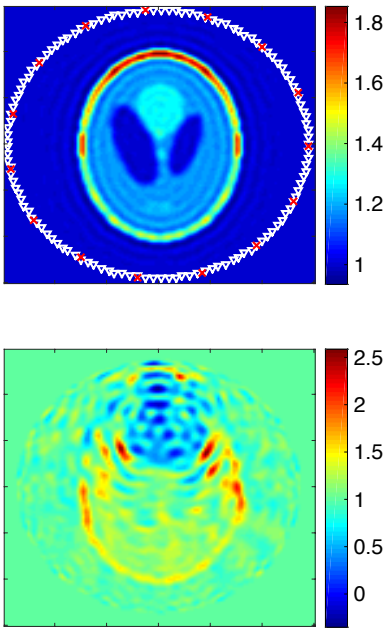


Figure 4: Shepp-Logan (reflectivity map). Recovered image for R2AS (top) and FWI (bottom) when starting from $\overline{m} = (1 - W(x))m_0$ background and $\tilde{m}_0 = 1$ initial perturbation. Receivers and sources are shown as white triangles and red crosses, respectively.

## REFERENCES


Burer, S., and R. D. Monteiro, 2003, A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization: Mathematical Programming, **95**, 329–357.

Chavent, G., and P. Lemonnier, 1974, Identification de la nonlinéarité d'une équation parabolique quasilinéaire: Applied mathematics and Optimization, **1**, 121–162.

Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: Geophysics, **51**, 1387–1403.

Haber, E., U. M. Ascher, and D. Oldenburg, 2000, On optimization techniques for solving nonlinear inverse problems: Inverse problems, **16**, 1263.

Lasserre, J. B., 2001, Global optimization with polynomials and the problem of moments: SIAM Journal on Optimization, **11**, 796–817.

Laurent, M., 2009, Sums of squares, moment matrices and optimization over polynomials, *in* Emerging applications of algebraic geometry: Springer, 157–270.

Lax, P. D., et al., 1957, Asymptotic solutions of oscillatory initial value problems: Duke Mathematical Journal, **24**, 627–646.

Lions, J. L., and E. Magenes, 1972, Non-homogeneous boundary value problems and applications: Springer Berlin.

Liu, D. C., and J. Nocedal, 1989, On the limited memory BFGS method for large scale optimization: Mathematical programming, **45**, 503–528.

Liu, Z., and N. Bleistein, 1995, Migration velocity analysis: Theory and an iterative algorithm: Geophysics, **60**, 142–153.

Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

Sava, P., and B. Biondi, 2004a, Wave-equation migration velocity analysis. I. Theory: Geophysical Prospecting, **52**, 593–606.

———, 2004b, Wave-equation migration velocity analysis. II. Subsalt imaging examples: Geophysical Prospecting, **52**, 607–623.

Shen, P., 2004, Wave equation migration velocity analysis by differential semblance optimization: PhD thesis, Rice University.

Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Symes, W., 2009, The seismic reflection inverse problem: Inverse problems, **25**, 123008.

Symes, W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

Symes, W. W., 1993, A differential semblance criterion for inversion of multioffset seismic reflection data: Journal of Geophysical Research: Solid Earth (1978–2012), **98**, 2061–2073.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

———, 2005, Inverse problem theory and methods for model parameter estimation: SIAM.

Tarantola, A., et al., 1984, The seismic reflection inverse problem: Inverse problems of acoustic and elastic waves, 104–181.

Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Reviews of Geophysics, **20**, 219–232.

van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667.

Versteeg, R., 1994, The Marmousi experience: Velocity model determination on a synthetic complex data set: The Leading Edge, **13**, 927–936.

Wu, R.-S., J. Luo, B. Wu, et al., 2013, Ultra-low-frequency information in seismic data and envelope inversion: Presented at the 2013 SEG Annual Meeting, Society of Exploration Geophysicists.