The diagonalator: inverse data space full waveform inversion

Peyman Poor Moghaddam, Delft University of Technology, Netherlands, and Wim A. Mulder, Shell Global Solutions International BV and Delft University of Technology, Netherlands*

SUMMARY

Full waveform inversion with the classic least-squares cost functional suffers from spurious local minima caused by cycle skipping in the absence of low frequencies in the seismic data. We present an alternative cost functional that is less sensitive to cycle skipping. It has the property of an annihilator, just like the functional used for velocity analysis with extended images based on subsurface shifts. For 2-D models with a line acquisition, the proposed functional applies a singular-value decomposition on the observed data and uses the eigenvectors to build a data panel that should be diagonal in the correct velocity model but has significant off-diagonal entries in the wrong model. By penalizing off-diagonal entries or maximizing values close to the main diagonal, the correct model should be found. We therefore call it the diagonalator.

A convexity test demonstrates the superiority of the proposed functional over the classic least-squares approach. We present initial synthetic data tests on a subset of a North Sea velocity model. The diagonalator performs better than least-squares data fitting in terms of resolving deeper events with full-bandwidth data. It also converges to an acceptable velocity model in the absence of low frequencies, when least-squares minimization fails.

INTRODUCTION

Local minima are a well-known problem for full waveform inversion (FWI). In Tarantola's minimum least-squares error framework (Tarantola and Valette, 1982), the synthetic waveforms in the starting model should match the observed data to within less than half a wavelength. Otherwise, cycle skipping may cause the minimization to end up in the nearest local minimum (Bunks et al., 1995; Chauris et al., 2008). To avoid this problem, a kinematically correct initial model is required. Alternatively, data with reliable low-frequency content may suffice to avoid the cycle-skipping problem (Plessix et al., 2010, e.g.).

The least-squares objective function (Tarantola and Valette, 1982) employs the squared L2-norm. Alternative objective functions are based on different norms, such the hybrid L1/L2 approach (Bube and Langan, 1997; Brossier et al., 2010) or the Huber norm (Huber, 1973; Guitton and Symes, 2003). Ha et al. (2009) applied the Huber norm for frequency-domain FWI and illustrated its robust behavior compared to the L2-norm when considering a dense frequency sampling during inversion. Although these objective functions are robust in the presence of large isolated and non-Gaussian errors, they basically suffer the same cycle-skipping problem that occurs in the L2-norm for higher-frequency data.

Moghaddam and Mulder (2012) introduced a new objective function that is less sensitive to cycle skipping. It was motivated by earlier work of van Leeuwen and Mulder (2008a), who proposed offset- or time-shifts in the data domain rather than in the subsurface domain. Subsurface space- or time-shifts enable velocity analysis in a wave-equation migration context (Faye and Jeannot, 1986; MacKay and Abma, 1992; Rickett and Sava, 2002; Sava and Fomel, 2006; Shen et al., 2003). Those methods may fail, however, when multiple energy is still present in the data (Mulder and ten Kroode, 2002; Mulder and van Leeuwen, 2008). The formulation in the data domain should maximize the correlation of the shifted data at zero lag. This approach can indeed handle multiples, at least in a NMO setting (van Leeuwen and Mulder, 2008b). The method does, however, suffer from cross-talk between events (van Leeuwen and Mulder, 2008a, 2010), requiring the use of a Gabor window to suppress contributions to the correlation from unrelated events. Also, its generalization to wave-equation methods requires a migration/demigration step to ensure that the observed and modeled that have similar events.

Our new method is inspired by the idea of the inverse data space (Berkhout, 2006). It introduces correlations between shifted data, as in the approach by van Leeuwen and Mulder (2008a), albeit in a way that is not transparent at this moment. We describe the method in the next section. We study its convexity, compared to the L1 and L2 error norms, by considering a horizontally layered model and computing the cost functionals for data in a sequence of models that are perturbations of the true one. Then, we present convergence results for two models, a simple horizontally layered model with three reflectors, starting from a velocity model that increases linearly with depth, and a more challenging marine model that is a subset of the North Sea model from Etgen and Regone (1998).

The new cost functional performs considerably better than least-squares data fitting, both for full bandwidth data, for which least-squares FWI should work, and for band-limited data with missing low frequencies, where least-squares FWI completely fails if the initial model is not kinematically correct in the wavelength range that corresponds to the absent frequencies.

METHOD

Consider a seismic survey with n_s shots and n_r receivers. We can arrange the data per frequency in a $n_r \times n_s$ data matrix $\mathbf{A}(\omega)$ of the form

$$\mathbf{A}(\omega) = \begin{pmatrix} r_{1,1}(\omega) & r_{1,2}(\omega) & \cdots & r_{1,n_s}(\omega) \\ r_{2,1}(\omega) & r_{2,2}(\omega) & \cdots & r_{2,n_s}(\omega) \\ \cdots & \cdots & \cdots & \cdots \\ r_{n_r,1}(\omega) & r_{n_r,2}(\omega) & \cdots & r_{n_r,n_s}(\omega) \end{pmatrix}, \quad (1)$$

with $r_{i,j}(\omega)$ the i^{th} receiver signal correspond to the j^{th} shot. If the receivers are not co-located for the different shots, as

DOI http://dx.doi.org/10.1190/segam2012-0210.1

common in a marine acquisition, some elements of the matrix $\mathbf{A}$ are set to zero (Berkhout, 2006). The singular-value decomposition (SVD) of the data matrix is $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{H}}$, where the unitary matrix $\mathbf{U}$ of size $n_r \times n_m$ contains the left eigenvectors as columns and $\mathbf{V}$ of size $n_s \times n_m$ the right eigenvectors as columns. The $n_m \times n_m$ diagonal matrix $\mathbf{S}$ consists of the singular values. The superscript $(\cdot)^{\mathrm{H}}$ denotes the conjugate transpose.

Given the SVD, $\mathbf{A}_{\mathrm{obs}}(\omega) = \mathbf{U}_{\mathrm{obs}}\mathbf{S}_{\mathrm{obs}}\mathbf{V}_{\mathrm{obs}}^{\mathrm{H}}$, of the observed data at an angular frequency ω, we expect $\mathbf{S} = \mathbf{U}_{\mathrm{obs}}^{\mathrm{H}}\mathbf{A}\mathbf{V}_{\mathrm{obs}}$ for computed data $\mathbf{A}$ to be diagonal if the velocity model is correct. Otherwise, there may be significant contributions on the off-diagonals. This observation motivates the choice of a cost functional that penalizes off-diagonal entries or maximizes entries close to and on the diagonal with, for instance, a gaussian weighting function. The corresponding cost functional is

$$J^{\mathrm{Diag}} = \sum_{\omega} \sum_{i=1}^{n_m} \sum_{j=1}^{n_m} w_{ij} |S_{ij}(\omega)|^2. \tag{2}$$

The weighting function w_{ij} is a function of $|i-j|$, the distance to the main diagonal. It should be decreasing with $|i-j|$ for maximization, for instance, with $w_{ij} = \exp(-\alpha|i-j|^2)$ and increasing for minimization, for instance, $w_{ij} = |i-j|^p$ with $p > 0$ or even $w_{ij} = \exp(|i-j|)$. We call this cost functional the diagonalator. It is an example of inverse data space full waveform inversion. Note that it requires multiple shots.

The matrix $\mathbf{S} = \mathbf{U}_{\mathrm{obs}}^{\mathrm{H}}\mathbf{A}\mathbf{V}_{\mathrm{obs}}$ in equation 2 depends on the current guess of the slowness model $\sigma(\mathbf{x}) = 1/v(\mathbf{x})$. The optimization problem becomes: minimize $J(\sigma)$ subject to the wave equation $(\omega^2\sigma^2 + \Delta)p_s + f_s = 0$, where $p_s(\omega, \mathbf{x})$ is the pressure wave field for shot number $s = 1, 2, \ldots, n_s$ as a function of frequency and position $\mathbf{x}$ and $f_s(\omega, \mathbf{x})$ represents the source term for shot s. The wave fields are collected into the matrix $\mathbf{A} = (D_1 p_1,\, D_2 p_2, \cdots,\, D_{n_s} p_{n_s})$, where D_s is the sampling or detection operator for shot s and its corresponding receivers. It maps the wavefield into a column vector of length n_r, containing zeros if necessary if receiver locations for subsequent shots do not fully coincide.

An annihilator (Symes, 2008) is an operator producing zero(s) when applied to the data or to an extended migration image in the correct velocity model and leads to a non-zero result otherwise. Although it is a concept that is defined for the high-frequency asymptotic approximation, it can have its value at finite frequencies. The method of Moghaddam and Mulder (2012) has the characteristics of an annihilator.

CONVEXITY

We consider a simple example to compare the behavior of the new functional to that the least-squares cost functional with the L1 or L2-norm in terms of convexity. We computed synthetic seismograms for multiple shots in the layered model shown in Figure 1a. We then perturbed the model by taking a weighted average with a model that has velocity increasing linearly with depth, shown in Figure 1b. The resulting model is $v(\alpha) = (1-\alpha)v_1 + \alpha v_2$, with $v_1(z)$ the layered and $v_2(z)$

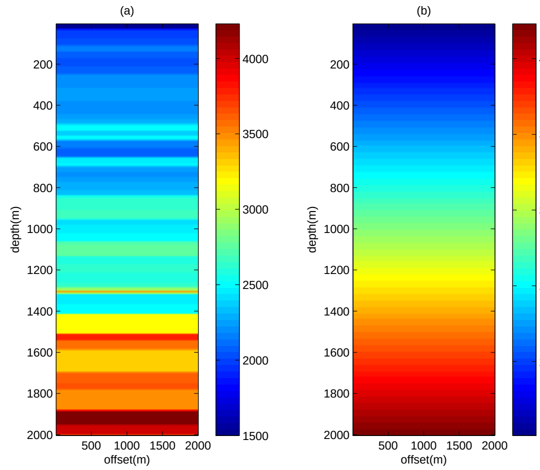

Figure 1: The velocity model for the convexity test is a weighted average of (a) the true model and (b) the linear model.

the linear velocity model. We computed synthetic data for 20 values of α ranging from -1 to 1 at frequencies between 8 and 30 Hz. Figure 2 shows the resulting values of the least-squares cost functional with either the L1- or L2-norm as well as the diagonalator values. Clearly, the diagonalator has a much wider basin of attraction than the least-squares cost functional and has better convexity properties. This implies that, with the diagonalator, a gradient-based minimization method will convergence towards the minimum for a starting model that can be further away from the correct model than with least-squares minimization.

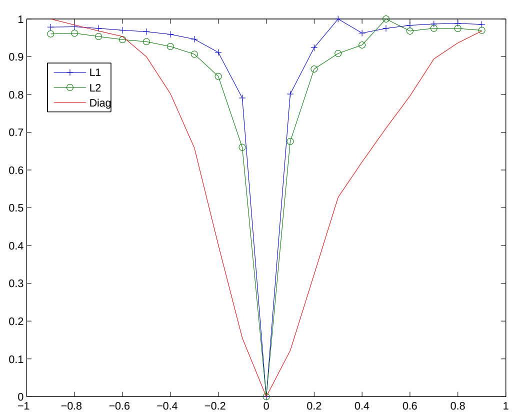

Figure 2: Comparison between the diagonalator, L1 and L2 objective functions.

INVERSION RESULTS

We tested the diagonalator against the traditional L2-norm FWI on a simple layered model shown in Figure 3a and the more complex North Sea model of Figure 5. For the minimization, we used a limited-memory BFGS method (Nocedal and

DOI http://dx.doi.org/10.1190/segam2012-0210.1

Wright, 2006). The gradient of the cost functionals was computed with the usual adjoint-state approach. The gradient computation for the diagonalator is similar to that for least-squares inversion but with the data residual replaced by a more complicated expression.

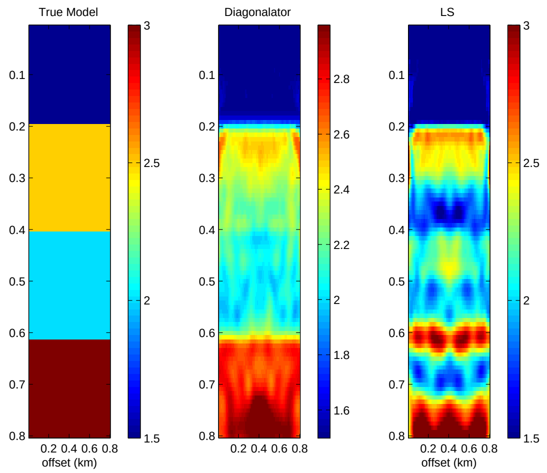

Figure 3: From left to right: (a) true model, (b) least-square inversion result, and (c) diagonalator reconstruction.

As a first test, we considered the very simple model in Figure 3a. We performed multi-scale inversion (Bunks et al., 1995; Ravaut et al., 2004) with data between 8 and 20 Hz, starting from an initial model with a linearly increasing velocity versus depth (not shown). Figure 3b shows the reconstructed model for the diagonalator, which is clearly superior to the least-squares inversion result with the L2-norm in Figure 3c. The latter displays the typical 'nonlinear migration' (Mulder and Plessix, 2008) character of FWI.

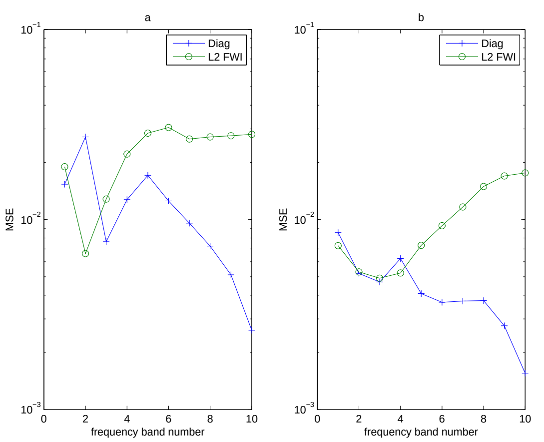

Figure 4: Average relative model error for 10 frequency bands of 2-Hz width between 0 an 20 Hz, using a smooth (a) or sharper (b) initial model.

To investigate in which frequency band the diagonalator is more effective, we divided the range from 0 to 20 Hz into 10 equally spaced bands of 2 Hz width, each band containing 4 different frequencies. The first band contains frequencies between 0 and 2 Hz, the second between 2 and 4 Hz, and so on. The starting model for the inversion was obtained by applying a spatial low-pass filter to the true model in Figure 3a. We prepared two versions, a smooth and a sharper one. For each band and each of the two starting models, we carried out 30 iterations with the L-BFGS minimization. Then, we determined the average relative model error, $\mathrm{MSE} = \mathrm{avg}\left([\hat{v}-v]^2/v^2\right)$, where $\mathrm{avg}(\cdot)$ denotes element-wise averaging and $\hat{v}$ is the inversion result. Figure 4a shows the resulting reconstruction errors for the smooth starting model. For the lower frequencies, the reconstructed model is smoother than the true one, so the error is expected to be larger than for the higher frequencies. The diagonalator clearly provides smaller errors in the higher frequencies, where least-squares inversion suffers from cycle skipping. Figure 4b displays similar results for a sharper starting model. Beyond band 4 (6–8 Hz), the least-squares cost functional starts to have problems.

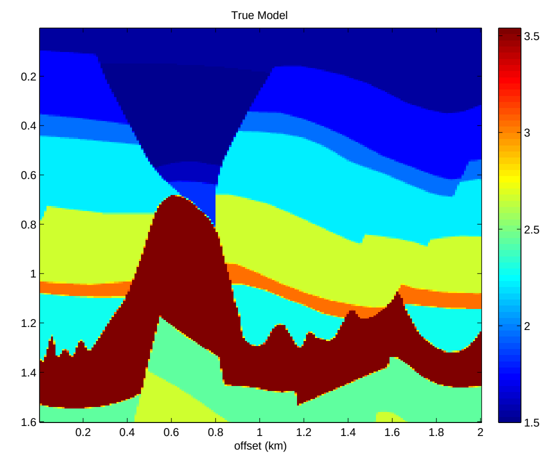

Figure 5: Subset of the North Sea velocity model.

We also tested our method on a part of North Sea velocity model, shown in Figure 5. The model contains steep dips as well as a salt structure. We first ran a full-bandwidth inversion with the multi-scale approach, using data between 0 and 20 Hz. In that case, we expect both the least-squares inversion and the diagonalator to produce an acceptable result. For the initial model, we took a horizontally layered model obtained by horizontally averaging the true model. Figure 6a shows the diagonalator result, which is somewhat better than the least-squares reconstruction in Figure 6b, especially around the steep dips.

We repeated the test on band-limited data with frequencies between 8 and 20 Hz, a typical range for vintage data sets. Figures 7a and 7b show the resulting reconstructions. The least-squares reconstruction fails, whereas the diagonalator is partially able to capture the structure of the model. Note that no smoothing or other type of regularizing penalty terms were applied during the minimization.

DOI http://dx.doi.org/10.1190/segam2012-0210.1

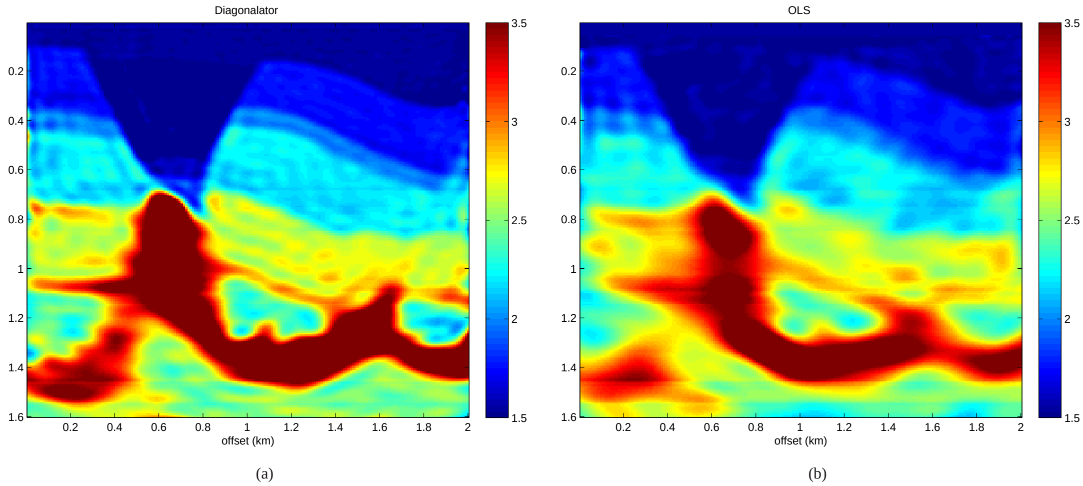

Figure 6: Full-bandwidth inversion test on a subset of the North Sea velocity model: (a) diagonalator reconstruction, (b) least-squares inversion result.

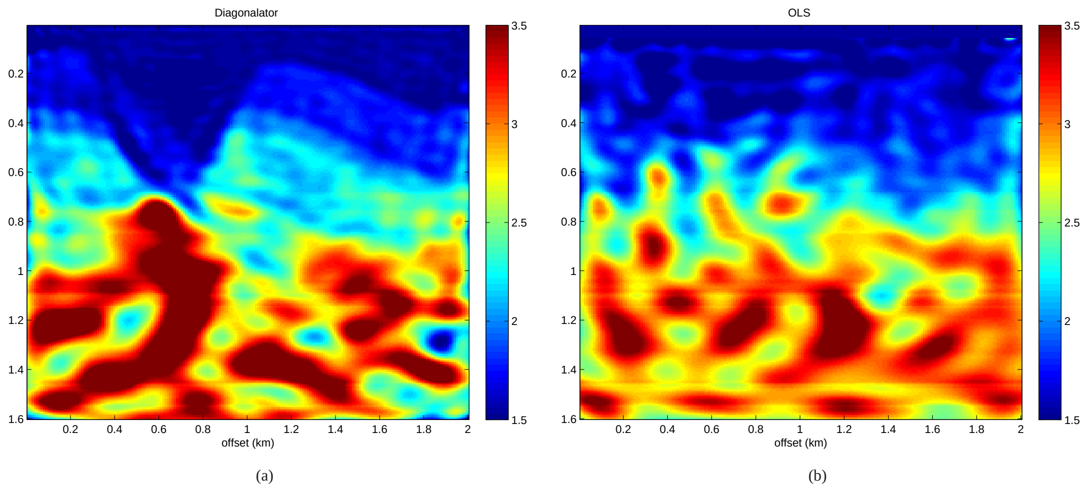

Figure 7: Band-limited inversion test on a subset of the North Sea velocity model: (a) diagonalator reconstruction, (b) least-squares inversion result.

CONCLUSIONS

We have proposed a new objective function for full waveform inversion, which we call the diagonalator. It is based on the singular-value decomposition of the observed data, ordered a in a specific way. If the corresponding matrices with eigenvectors are applied to the modeled data, the resulting matrix should be diagonal in the correct velocity model. The cost functional penalizes off-diagonal entries. We have demonstrated that the objective function is superior to the classic least-squares cost functional, either with the L1- or L2-norm, in terms of convexity and accuracy of the reconstructed model.

ACKNOWLEDGEMENTS

The first author received financial support from Shell GameChanger.

DOI http://dx.doi.org/10.1190/segam2012-0210.1

http://dx.doi.org/10.1190/segam2012-0210.1

EDITED REFERENCES

Note: This reference list is a copy-edited version of the reference list submitted by the author. Reference lists for the 2012 SEG Technical Program Expanded Abstracts have been copy edited 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

Berkhout, A. J., 2006, Seismic processing in the inverse data space: Geophysics, **71**, no. 4, A29–A33, doi:10.1190/1.2217727.

Brossier, R., S. Operto, and J. Virieux, 2010, Which data residual norm for robust elastic frequency-domain full waveform inversion?: Geophysics, **75**, no. 3, R37–R46, doi: 10.1190/1.3379323.

Bube, K. P., and R. T. Langan, 1997, Hybrid [script-1][sup 1]/[script-1][sup 2] minimization with applications to tomography: Geophysics, **62**, no. 4, 1183–1195, doi: 10.1190/1.1444219.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, no. 5, 1457–1473, doi: 10.1190/1.1443880.

Chauris, H., M. Noble, and C. Taillandier, 2008, What initial velocity model do we need for full waveform inversion?: 70th Conference and Exhibition, EAGE, Extended Abstracts.

Etgen, J., and C. Regone, 1998, Strike shooting, dip shooting, widepatch shooting: Does prestack depth migration care? A model study: 68th Annual International Meeting, SEG, Expanded Abstracts, 66–69, doi: 10.1190/1.1820541.

Faye, J., and J. Jeannot, 1986, Prestack migration velocities from focusing depth analysis: 56th Annual International Meeting, SEG, Expanded Abstracts, 438–440.

Guitton, A., and W. W. Symes, 2003, Robust inversion of seismic data using the Huber norm: Geophysics, **68**, no. 4, 1310–1319, doi: 10.1190/1.1598124.

Ha, T., W. Chung, and C. Shin, 2009, Waveform inversion using a back-propagation algorithm and a Huber function norm: Geophysics, **74**, no. 3, R15–R24, doi: 10.1190/1.3112572.

Huber, P. J., 1973, Robust regression: Asymptotics, conjectures, and Monte Carlo: The Annals of Statistics, **1**, no. 5, 799–821.

MacKay, S., and R. Abma, 1992, Imaging and velocity analysis with depth-focusing analysis: Geophysics, **57**, no. 12, 1608–1622, doi: 10.1190/1.144322.

Moghaddam, P. P., and W. A. Mulder, 2012, The diagonalator, an alternative cost functional for wave-equation inversion: 74th Conference and Exhibition, EAGE, Extended Abstracts.

Mulder, W. A., and R.-E. Plessix, 2008, Exploring some issues in acoustic full waveform inversion: Geophysical Prospecting, **56**, no. 6, doi: 10.1111/j.1365-2478.2008.00708.x.

Mulder, W. A., and A. P. E. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: Geophysics, **67**, no. 4, 1184–1191, doi: 10.1190/1.1500380

Mulder, W. A., and T. van Leeuwen, 2008, Automatic migration velocity analysis and multiples: 78th Annual International Meeting, SEG, Expanded Abstracts, 3128–3132, doi:10.1190/1.3063996

Nocedal, J., and S. Wright, 2006, Numerical optimization, 2nd ed.: Springer Series in Operations Research and Financial Engineering: Springer-Verlag.

Plessix, R.-E., G. Baeten, J. W. de Maag, M. Klaassen, Z. Rujie, and T. Zhifei, 2010, Application of acoustic full waveform inversion to a low-frequency large-offset land data set: 80th Annual International Meeting, SEG, Expanded Abstracts, 930–934.

Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero, and P. Dell'Aversana, 2004, Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: Application to a thrust belt: Geophysical Journal International, **159**, no. 3, 1032–1056, doi: 10.1111/j.1365-246X.2004.02442.x.

Rickett, J. E., and P. C. Sava, 2002, Offset and angle-domain common image-point gathers for shot-profile migration: Geophysics, **67**, no. 3, 883–889, doi: 10.1190/1.1484531.

Sava, P., and S. Fomel, 2006, Time-shift imaging condition in seismic migration: Geophysics, **71**, no. 6, S200–S217, doi: 10.1190/1.2338824.

Shen, P., W. W. Symes, and C. C. Stolk, 2003, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, SEG, Expanded Abstracts, 2132–2135, doi: 10.1190/1.1817759.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, no. 6, 765–790, doi: 10.1111/j.1365-2478.2008.00698.x.

Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least square criterion: Reviews of Geophysics, **20**, no. 2, 219–232, doi: 10.1029/RG020i002p00219.

van Leeuwen, T., and W. A. Mulder, 2008a, Velocity analysis based on data correlation: Geophysical Prospecting, 56, no. 6, 791–803, doi: 10.1111/j.1365-2478.2008.00704.x

van Leeuwen, T., and W. A. Mulder, 2008b, Velocity analysis with multiples – NMO modeling for layered velocity structures: SEG Technical Program Expanded Abstracts, 27, 1925–1929, doi: 10.1190/1.3059275

van Leeuwen, T., and W. A. Mulder, 2010, A comparison of seismic velocity inversion methods for layered acoustics: Inverse Problems, 26, no. 1, 015008, doi:10.1088/0266-5611/26/1/015008

DOI http://dx.doi.org/10.1190/segam2012-0210.1