Sparsity-promoting migration from surface-related multiples

Tim T.Y. Lin, Ning Tu, and Felix J. Herrmann, University of British Columbia, EOS*

SUMMARY

Seismic imaging typically begins with the removal of multiple energy in the data, out of fear that it may introduce erroneous structure. However, seismic multiples have effectively seen more of the earth's structure, and if treated correctly can potential supply more information to a seismic image compared to primaries. Past approaches to accomplish this leave ample room for improvement; they either require extensive modification to standard migration techniques, rely too much on prior information, require extensive pre-processing, or resort to full-waveform inversion. We take some valuable lessons from these efforts and present a new approach balanced in terms of ease of implementation, robustness, efficiency and well-posedness, involving a sparsity-promoting inversion procedure using standard Born migration and a data-driven multiple modeling approach based on the focal transform.

INTRODUCTION

There are many arguments in favour of generating seismic images using multiples, provided that one can actually do it. As multiples take roundtrips betwixt the earth's subsurface it can pierce beyond the first reflection and see more structural information than the primary events that gave rise to it. Outside of the zero-offset context, the multiples also provides wider illumination angles than primaries. Furthermore, after normalization of the energy difference, multiples are by definition more sensitive to velocity and reflectivity perturbations than primaries, and under a pure informational context one can emphatically argue that it's consequently a more crucial ally to the interest of deducing structural models of the subsurface. All this hinges, of course, on the existence and feasibility of proper methods that can extract this additional information, and manages to avoid becoming confused by it.

Multiple reflections are notoriously fussy about how it is treated in the context of seismic imaging, which is why the traditional approach to migrating multiples always starts with its elimination. When sent directly into straightforward migration routines devised under the assumptions of smooth background velocity models and linearized perturbations, multiples end up as false, spatially displaced copies of pre-existing events that conflict with those corresponding to other primary reflections. Only in the past fifteen years have successful attempts been made to actually use them meaningfully in the imaging step.

At least in terms of surface-related multiples, there are three emergent schools of methods to incorporate them as a part of imaging (Verschuur, 2006). Firstly we can modify existing migration techniques to bring multiples to the table. One can, of course, model them directly by including free surfaces in the background velocity model so the Green's functions are at least partially capable of describing the multiples (Reiter et al., 1991). To do this one would have to give up the notion of smooth background velocities, which amongst other drawbacks also poses fairly stringent requirements on the calculation of the Green's functions. It also raises the question of how the validity of linear perturbation assumption becomes affected by the burden of explaining multiple events. If one were to take advantage of the fact that primary events are sources of multiple events then an alternative scheme similar to reverse-time migration can be derived for the one-way setting (Berkhout and Vershuur, 1994; Muijs et al., 2005; Guitton, 2002; Sheng, 2001). Similar derivation for the two-way wave equation setting exists but requires accurate knowledge of the boundary conditions of the model before imaging (Youn and Zhou, 2001).

One can also follow the move towards waveform inversion. We note that our problem on hand is a classic example of the original motivations of Tarantola (1984) in his proposition to unite all of seismic signal processing into one joint non-linear inversion process. Doing this successfully requires the ability to fully handle the complexities and difficulties associated with full waveform inversion, making it an all-or-nothing approach. Although FWI is arguably the future of seismic imaging, for now it is likely to be out of the question if a migration-based imaging is all one can afford.

Finally there is the approach that handles the primary-multiples relationship in a more data-driven way, namely that of using focal transforms (Berkhout and Verschuur, 2003). This method has roots in interferometry principles, and relies on the fact that a non-stationary convolution between the primary impulse response and the upgoing wavefield reproduces the multiples itself. By applying the pseudoinverse of the primary impulse response to the total data (followed by suitable time-shifting and windowing) one can transform the first multiples into primaries, while mapping the original primaries back to the source. Migration is then done on the newly crafted multiples after applying more traditional de-multiple techniques. While requiring little prior knowledge, this technique must be carried out over several iterative steps that may introduce compounding errors at each step.

Here we present a method to image multiples that takes advantageous elements from all these approaches. We start with the traditional least-squares migration using smooth background velocities and linearized reflectivity. The multiple information is then added using an evolution of the focal transform, which is a recent method in primary estimation proposed by van Groenestijn and Verschuur (2009). To tie it all up and bring stability to the process, we combine the two operation in a joint sparsity-regulated inversion process in the spirit of Tarantola, with the important distinction that our inversion problem is less ambitious in scale and linear in nature. We believe our approach takes the right compromise between the different methods to arrive at formulation balanced in terms of ease of implementation, robustness, efficiency and well-posedness.

The following sections begin with separate background discussions, respectively, on a sparsity-promoting primary estimation method and the Born-scattering based least-squares migration. We will note how both methods combine to naturally lead to a sparsity-promoting linear inversion that images from the total upgoing data. This will in turn lead to a discussion on a way to efficiently carry out the inversion using modern optimization techniques based on pure gradient steps that minimizes ℓ_1-norm objectives. Finally, we will show numerical examples of this inversion scheme for synthetic data.

TRANSFORMING PRIMARIES INTO MULTIPLES

Currently in seismic literature there are some interesting non-imaging techniques similarly aimed at extracting additional information from multiples. One that stands out among them is a method called Estimation of Primaries from Sparse Inversion (EPSI), which is particularly of interest to us (van Groenestijn and Verschuur, 2009). EPSI is essentially a scheme that simultaneous solves for the (surface-related) primary impulse-response and the source-wavelet signature, by an Amundsen-type inversion from an operator that models total upgoing field data from the primary impulse-response. This modeling operator is essentially the focal transform run backwards, making EPSI likewise a data-driven method.

The EPSI approach has several benefits: Firstly, due to the wavefield inversion context, EPSI will attempt to minimize the energy of the modeling residual, taken over the entire dataset. This is contrary to prediction-subtraction based techniques, where the minimization is over the energy of the data minus the predictions, possibly corrupting primary energy in the process. Furthermore, the data-driven approach is agnostic to an inherent assumption of a certain periodicity range, and is considered more robust compared to wave-equation based techniques. Thanks to a formulation that allows simultaneous inversion of various variables, it is also possible for EPSI to infer missing-near offset traces from the multiple data. The spirit of extracting information from multiples is epitomized by this last point; starting with just the multiple data itself, we can obtain estimates of the source wavelet and any missing near-offset traces, armed with nothing more than a convolution-based relation of the primaries and the multiple wavefield.

EPSI in an optimization context

The EPSI modeling operator is formulated from the same underlying principle as the focal transform, namely the non-stationary transform relation, which relates the primary impulse response to the total up-going wavefield that includes the source signature and surface-related multiples. We can describe this in a discretized setting if we introduce the notion of monochromatic slices of wavefields arranged into a matrix that have columns representing common shot gathers, similar to the detail-hiding notation of Berkhout and Pao (1982), such that the matrix multiplication of two hatted wavefield quantities become non-stationary convolutions in the time domain. With this notation, we can write

$$\hat{\mathbf{P}} = \hat{\mathbf{G}}(\hat{\mathbf{Q}} + \mathbf{R}\hat{\mathbf{P}}), \quad (1)$$

where $\mathbf{G}$ represents the primary impulse response, $\mathbf{P}$ the total up-going wavefield, and $\mathbf{Q}$ a (possibly shot-dependent) source signature function. Hatted quantities represent monochromatic variables. $\mathbf{R}$ is the reflection coefficient at the surface that is approximated to $-\mathbf{I}$ for the rest of this text.

As foreshadowed, EPSI involves solving an inverse problem. In order facilitate further discussion of this we now try to mathematically reformulate EPSI to be consistent with the usual notations of general optimization problems, i.e. the canonical form of solving for an unknown vector quantity $\mathbf{x}$ from a vector observation $\mathbf{b}$, using the relationship $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A}$ is some linear operator. This is usually accomplished by minimizing some objective $\|\mathbf{b} - \mathbf{Ax}\|$ plus additional regularization terms on $\mathbf{x}$.

By introducing the convention of vectorized wavefields in lower case, e.g. $\mathbf{p} = \mathrm{vec}(\mathbf{P})$, we can then express Eq. 1 in terms of a linear operator $\mathbf{M}$ acting on vectorized primary impulse response $\mathbf{g}$:

$$\mathbf{Mg} := \mathscr{F}_\mathrm{t}^* \mathrm{BlockDiag}_f[(\hat{\mathbf{Q}} - \hat{\mathbf{P}}^-)^* \otimes \mathbf{I}]\mathscr{F}_\mathrm{t}\mathbf{g} = \mathbf{p}, \quad (2)$$

where the block diagonal elements varies over frequency. $\otimes$ defines a Kronecker product that, in this case, helps reformulate matrix multiplication into matrix-vector products. $\mathscr{F}_\mathrm{t}$ is a Fourier transform in the time axis that also organized the different monochromatic wavefields in a vectorized manner, such that $\mathscr{F}_\mathrm{t}\mathbf{g} = \hat{\mathbf{g}} := [\hat{\mathbf{g}}_{f1}, \hat{\mathbf{g}}_{f2}, ..., \hat{\mathbf{g}}_{fn}]^T$, while the adjoint operation $\mathscr{F}_\mathrm{t}^*$ on the left brings the wavefield back to the time domain. Note that $\mathbf{M}$ is a simple linear operator that depends both on a source signature estimate $\mathbf{Q}$ and the (recorded) up going wavefield $\mathbf{P}$. In most cases $\mathbf{Q}$ is an unknown quantity that must be explicitly inverted for at the same time as $\mathbf{g}$. However, in this abstract we are mainly interested in the interplay between the EPSI modeling operator and a migration operator, and in order to focus on that aspect we will assume from now on $\mathbf{Q}$ to be known to an appropriate accuracy.

EPSI states that a reasonable estimate of the primary impulse response can be obtained by a steepest-descent inversion process. The gradients of the objective function $f(\mathbf{g}) = \|\mathbf{p} - \mathbf{Mg}\|_2^2$, is evaluated at $\tilde{\mathbf{g}}$ (an estimate on $\mathbf{g}$) according to

$$\nabla f|_{\tilde{\mathbf{g}}} = 2\,(\mathbf{p} - \mathbf{M}\tilde{\mathbf{g}})\,\mathbf{M}^*. \quad (3)$$

A sparse update $\delta\tilde{\mathbf{g}}$ on $\tilde{\mathbf{g}}$ is then obtained by picking the τ-th largest elements of the gradient and setting the rest to zero, followed by a scaling factor determined by a simple line search. The next update will then be calculated on the gradient of $(\tilde{\mathbf{g}} + \delta\tilde{\mathbf{g}})$. The process is repeatedly carried out until a desired image of the primary is formed. The typical number of iterations required for convergence is claimed by the authors to be on the neighborhood of 100.

Since steepest decent methods belong to the class of gradient methods in optimization, it is beneficial to look at EPSI in terms of an optimization instance. The goal of EPSI is to solve an instance of the following *non-convex* optimization problem:

$$\tilde{\mathbf{g}} = \underset{\mathbf{g}}{\mathrm{argmin}}\, \|\mathbf{p} - \mathbf{Mg}\|_2 \quad \text{subject to} \quad \|\mathbf{g}\|_0 \leq k\tau, \quad (4)$$

where k is the number of iterations taken in the EPSI process. The ℓ_0 pseudonorm $\|\mathbf{g}\|_0$ measures the *cardinality* – the number of non-zero elements, or in other words the *sparsity* – in $\mathbf{g}$.

Dealing with an ℓ_0 measure in optimization is famously difficult, since uniqueness and existence of a solution cannot be proved in general and no method outside of combinatorial searches are guarenteed to yield the optimal solution. All these problems are inherent in the original formulation of EPSI as well. EPSI works around this limitation by severely limiting the size of the feasible set at every iteration, such as imposing constraint $\|\delta\tilde{\mathbf{g}}\|_0 = \tau$, so that the update making the largest possible progress to the minimization objective in Eq. 4 can be found by simple searching over the whole set, i.e. zeroing everything in the gradient except the τ-th largest elements. Note that final solution by definition will remain feasible for the original constraint $\|\mathbf{g}\|_0 < k\tau$.

Instead of using various regularizations on the update to stabilize the ill-posed EPSI problem, one can postulate a similar but much more tractable problem by performing a convexification (Lin and Herrmann, 2010). A very well-known strategy when tackling cardinality-constrained problems is to replace the ℓ_0 term with the ℓ_1 norm $\|\mathbf{g}\|_1$, the sum over the element-wise absolute value of $\mathbf{g}$, leading to

$$\tilde{\mathbf{g}} = \underset{\mathbf{g}}{\operatorname{argmin}} \|\mathbf{p} - \mathbf{M}\mathbf{g}\|_2 \quad \text{subject to} \quad \|\mathbf{g}\|_1 \leq \tau, \qquad (5)$$

where the τ in this expression is overloaded to be an ℓ_1 norm constraint on the model. Since taking an ℓ_1-norm is a convex function, this formulation of EPSI is also convex. Convexity is a desirable property, because convex problems in general are stable with no local minima in the objective function. A useful consequence of a convex EPSI formulation is that we are allowed to exploit a well known duality result (van den Berg and Friedlander, 2008) to associate a certain value of τ in eq. 5 with a certain value of σ in the following problem:

$$\tilde{\mathbf{g}} = \underset{\mathbf{g}}{\operatorname{argmin}} \|\mathbf{g}\|_1 \quad \text{subject to} \quad \|\mathbf{p} - \mathbf{M}\mathbf{g}\|_2 \leq \sigma, \qquad (6)$$

such that both problems lead to the same solution. Here σ is seen as the residual between the recorded data and the total up-going wavefield predicted by the estimated primaries, which is highly linked to the noise level of the shot record. Due to its physical significance, a good estimate for σ should be more easily determined compared to τ.

SPARSITY-PROMOTING MIGRATION

EPSI is an ideal preprocessing step for migration-based imaging because it removes both multiple information and source signatures in one fell swoop. Because these difficult to deal with artifacts are removed, one can have a lot of freedom in choosing a suitably simple migration routine, for example the venerable Kirchoff migration. However, we will now point out that with a suitable choice of migration methods we might get an additional boost from EPSI.

Our starting is a migration technique that similarly requires an inversion process: least-squares migration from the linearized Born scattering operator. We are interested in the linearized reflectivity $\delta\mathbf{m}$, which under the linearized Born operator $\mathbf{J}$ produces a linearized scattered wavefield $\mathbf{b}$. Now assuming that our primary impulse response $\mathbf{g} \approx \mathbf{b}$ is a good approximation for the primary impulse response, we may image the reflectivity by solving the following regularized least-squares problem:

$$\delta\tilde{\mathbf{m}} = \underset{\delta\mathbf{m}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{g} - \mathbf{J}\delta\mathbf{m}\|_2^2 + \lambda\|\delta\mathbf{m}\|_2^2, \qquad (7)$$

for some ℓ_2-norm regularization parameter λ.

By entering once again into the inversion context we can then appeal to an old observations in seismic inversion: subsurface reflection events rarely occur in a Gaussian-distribution fashion, which the ℓ_2-norm regularization assumes. Rather, the singularities distribution should be assumed to have Laplacian priors, which calls instead for an ℓ_1-norm regularization term (Menke, 1989). We can rewrite this ℓ_1-norm regularization term in constraint form similar to eq. 5, and use the same duality result used to derive eq. 6 to arrive at the following ℓ_1-regularized formulation:

$$\delta\tilde{\mathbf{m}} = \underset{\delta\mathbf{m}}{\operatorname{argmin}} \|\delta\mathbf{m}\|_1 \quad \text{subject to} \quad \|\mathbf{g} - \mathbf{J}\delta\mathbf{m}\|_2 \leq \sigma, \qquad (8)$$

where we override σ from eq. 6.

Equation 8 hints to where we can use the EPSI modeling operator to our benefit. The trick is to setup the operator $\mathbf{M}$ as a type of left pre-conditioner for $\mathbf{J}$. Together with eq. 2 we can finally arrive at

$$\delta\tilde{\mathbf{m}} = \underset{\delta\mathbf{m}}{\operatorname{argmin}} \|\delta\mathbf{m}\|_1 \quad \text{subject to} \quad \|\mathbf{p} - \mathbf{M}\mathbf{J}\delta\mathbf{m}\|_2 \leq \sigma, \qquad (9)$$

where σ is again overridden for the last time, and $\mathbf{p}$ is again the total upgoing wavefield. Due to the close link of the ℓ_1-norm to the degree of sparsity (ℓ_1-norm minimizing is often called sparsity promoting), this optimization problem now poses a physically intuitive question: Given a particular confidence/noise-level in our total recorded upgoing data, find the sparsest set of linearized reflectivity that explains this data wavefield, *including* the surface-related multiple reflections.

We are not certain whether using $\mathbf{M}$ as a pre-conditioner actually improves the numerical conditioning of the system. Instead, we make an observation of the two very significant advantages of this method:

1. Using nothing more than a purely data driven model for the surface-related multiples, and a relatively standard migration technique, we arrived at a straightforward way to image from data “contaminated” with multiples and the source signature.
2. The ℓ_1-norm minimization form of inverting $\mathbf{M}$ and $\mathbf{J}$ (eqns. 6 and 8) are derived from different arguments, but both with the goal of inversion stability and well-posedness. We therefore expect our inversion formulation in eq. 9 to behave in a stable fashion as well.
3. Because it takes into account multiple information, $\mathbf{MJ}$ should be more sensitive to image model perturbations $\delta\mathbf{m}$ than $\mathbf{J}$ alone. Provided that the inversion is carried out stably (see point 2), eq. 9 should recover more structural information than eq. 8, converge faster, or both.

To backup the claims made in points (2) and (3), we devise a numerical example using linearized data generated from a synthetic salt dome marine model with a relatively shallow water depth, preceded by a short discussion of feasible and affordable methods for ℓ_1-norm minimization.

NUMERICAL EXAMPLES

Efficient gradient methods for ℓ_1 problems

We now address the issue of realistically and efficiently carrying out the ℓ_1-norm minimization problems posed in eqns. 6, 8, and 9. Gradient-based methods, while maintaining the comfortable form of steepest-descent, often perform poorly due to the inherent non-differentiability of the ℓ_1 norm. Traditional approaches to ℓ_1 type problems almost always involve recasting it to an equivalent linear programming problem, solved with either the simplex method or more modernly a interior-point type method. Both of these approaches involves dealing directly with the linear operator in question (e.g., **MJ** for our case), which is prohibitively expensive with the size of these operators involved in imaging form real-world data.

Thankfully, due to the recent interests in compressive sensing and sparsity-promoting algorithms in general (mostly in other fields such as medical imaging and machine learning), some practical methods applicable to very large data have lately emerged. This includes several gradient-based methods, which work around the non-differentiability in differing and clever ways without resorting to subgradient-type (often slow to converge) approaches. Of interest to us is the SPGℓ_1 method (van den Berg and Friedlander, 2008), which uses Pareto root-finding schemes to determine a suitable τ that solves problems of type eq. 5 to give an identical solution to solving problems of type eq. 6 for a certain σ. We note that this method requires the least amount of gradient evaluations out of all currently available methods (Becker et al., 2009), while rarely requiring line searches. Anecdotally we can claim that solving eq. 6 for sizable wavefields rarely requires over 100 gradient evaluations.

Imaging from synthetic marine model

We present in Figure 1 some imaging results from linearized data. The background velocity model is a synthetic salt dome shown in **1a** with the first layer being a 200m deep water layer, and the receiver and shot spacing is 10m. A reference linearized reflectivity $\delta\mathbf{m}$ is obtained by taking the derivative in the depth direction, and the background velocity is obtain with a 10-point averaging filter. Time domain FD forward modeling is done for the $\mathbf{P}$ term used in $\mathbf{M}$ with $dt = 6$ms for 1.5s, using a Ricker source wavelet with maximum frequency at 60Hz and dominant frequency near 30Hz. Linear data generation and the subsequent modeling step are done in the frequency domain using 15 randomly selected frequencies. The inversion is carried out using SPGℓ_1. **1b** shows the solution of eq. 8 using $\mathbf{J}$ to generate $\mathbf{b}$ from $\delta\mathbf{m}$. **1c** shows the solution of eq. 9 using $\mathbf{MJ}$ to generate $\mathbf{p}$ from $\delta\mathbf{m}$. **1d** shows the result of solving eq. 8 using the $\mathbf{p}$ from **1c** as input. All the above are solved using 140 gradient evaluations. **1e** and **1f** show the imaging results of **1b** and **1c** using a quarter of the evaluation costs (30 gradient evaluations).

ACKNOWLEDGEMENTS

We would like to thank Gert Jan van Groenestijn and Eric Verschuur at DELPHI for discussions and help on EPSI. This work was financed through the SINBAD II project with support from BG, BP, Chevron, Petrobras, and WesternGeco, and was partly supported by a NSERC Discovery Grant (22R81254) and CRD Grant DNOISE (334810-05).

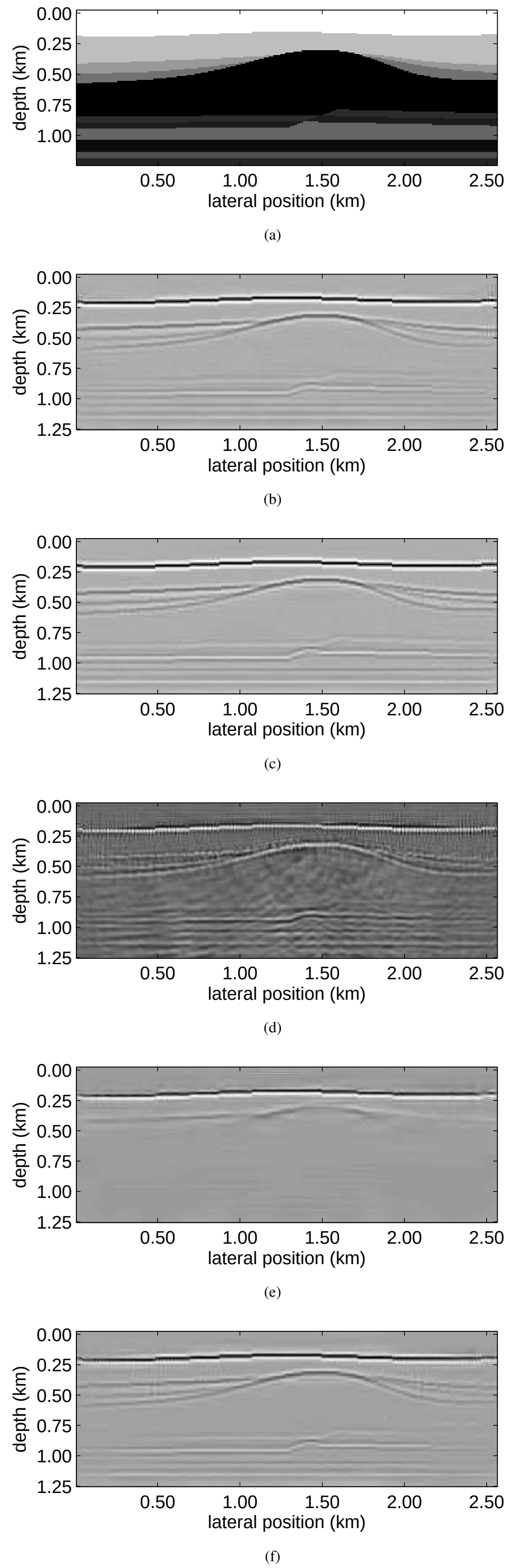

Figure 1: **(a)** Reference velocity model, **(b)** imaging from linearized data with straightforward migration, **(c)** imaging from linearized data with multiples using our scheme, **(d)** imaging from linearized data with multiples using straightforward migration, note the false events inside the model and the generally lower image quality.**(e)** and **(f)** are respectively **(b)** and **(c)** using a quarter of the evaluation cost.

REFERENCES

Becker, S., J. Bobin, and E. Candès, 2009, Nesta: A fast and accurate first-order method for sparse recovery: Technical report, California Institute of Technology.

Berkhout, A. J., and Y.-H. Pao, 1982, Seismic migration—imaging of acoustic energy by wave field extrapolation: Journal of Applied Mechanics, **49**, 682–683.

Berkhout, A. J., and D. J. Verschuur, 2003, Transformation of multiples into primary reflections: SEG Technical Program Expanded Abstracts, **22**, 1925–1928.

Berkhout, A. J., and D. J. Vershuur, 1994, Multiple technology: Part 2, migration of multiple reflections: SEG Technical Program Expanded Abstracts, **13**, 1497–1500.

Guitton, A., 2002, Shot-profile migration of multiple reflections: SEG Technical Program Expanded Abstracts, **21**, 1296–1299.

Lin, T., and F. J. Herrmann, 2010, Stabilized estimation of primaries via sparse inversion: Presented at the 72nd EAGE Conference & Exhibition.

Menke, W., 1989, Geophysical data analysis: Academic Press.

Muijs, R., K. Holliger, and J. O. A. Robertsson, 2005, Prestack depth migration of primary and surface-related multiple reflections: SEG Technical Program Expanded Abstracts, **24**, 2107–2110.

Reiter, E. C., M. N. Toksöz, T. H. Keho, and G. M. Purdy, 1991, Imaging with deep-water multiples: Geophysics, **56**, 1081–1086.

Sheng, J., 2001, Migrating multiples and primaries in cdp data by crosscorrelogram migration: SEG Technical Program Expanded Abstracts, **20**, 1297–1300.

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

van den Berg, E., and M. P. Friedlander, 2008, Probing the pareto frontier for basis pursuit solutions: SIAM Journal on Scientific Computing, **31**, 890–912.

van Groenestijn, G. J. A., and D. J. Verschuur, 2009, Estimation of primaries and near-offset reconstruction by sparse inversion: Marine data applications: Geophysics, **74**, R119–R128.

Verschuur, D. J., 2006, Seismic multiple removal techniques: Past, present and future: EAGE Publications BV.

Youn, O. K., and H. W. Zhou, 2001, Depth imaging with multiples: Geophysics, **66**, 246–255.