Imaging with multiples accelerated by message passing

Ning Tu and Felix J. Herrmann

SUMMARY

With the growing realization that multiples can provide valuable information, there is a paradigm shift from removing them to using them. For instance, primary estimation by sparse inversion has demonstrated its superiority over surface-related multiple removal in many aspects. Inspired by this shift, we propose a method to image directly from the total up-going wavefield, including surface-related multiples, by sparse inversion. To address the high computational cost associated with this method, we propose to speed up the inversion by having the wave-equation solver carry out the multi-dimensional convolutions implicitly and cheaply by randomized subsampling. We improve the overall performance of this algorithm by selecting new independent copies of the randomized modeling operator, which leads to a cancellation of correlations that hamper the speed of convergence of the solver. We show the merits of our approach on a number of examples.

INTRODUCTION

Conventionally, surface-related multiples are removed before migration to avoid the introduction of phantom reflectors into seismic images (Tu et al., 2011). While this removal achieves this goal, there is a growing realization that multiples provide useful additional information. Therefore, we propose a method that uses the wider illumination of multiples during imaging instead of removing them (Verschuur, 2006).

Accurate imaging with multiples hinges on two factors, namely the incorporation of the surface multiple-prediction operator in linearized-Born scattering and inversion to avoid artifacts arising from non-causal cross-correlations (Lin et al., 2010; Whitmore et al., 2010; Verschuur, 2011; Liu et al., 2011).

In our work, we also rely on incorporating the relationship between up- and down-going waves at the surface into the linearization of the wave equation. This amounts to replacing impulsive point sources by an areal source given by the total down-going wavefield. As reported in Tu and Herrmann (2012), this leads to a computational efficient formulation because we avoid calculating multi-dimensional convolutions explicitly. This is important because our method relies on sparse inversion (Lin et al., 2010), which requires multiple iterations.

Unfortunately, the cost of wavefield simulations are prohibitively large because they increase with the number of sources. Motivated by recent work on random source encoding, we reduce the required number of sources experiments by selecting subsets of frequencies and simultaneous sources (Tu and Herrmann, 2012). This reduction not only decreases the simulation costs but also reduces memory use and IO. To improve the rate of convergence of the model error, we follow recent insights from compressive sensing(Donoho, 2006), and in particular from message passing (Donoho et al., 2009), which relies on selecting independent randomized source experiments throughout the optimization.

The outline of this abstract is as follows. First, we incorporate the relation between up- and down-going waves into the wave equation followed by a proposal to increase the efficiency by using simultaneous sources combined with curvelet-domain sparsity promotion. We conclude by demonstrating the uplift of drawing independent experiments that break correlations between the model iterate and the subsampling operator.

INCORPORATING THE FREE SURFACE

To facilitate further discussion, we first introduce three wavefields: the total up-going wavefield $\mathbf{P}$ that contains both primaries and multiples, the surface free Green's function $\mathbf{G}$, and the point source wavefield $\mathbf{Q}$. All wavefields are considered in the *frequency* domain, the dimensions of which are receiver positions, source positions, and frequencies. Throughout this abstract, we assume $\mathbf{Q}_{\omega_i}$ to be $q(\omega_i)\mathbf{I}$, where $\mathbf{I}$ is the identity matrix implying uniform source directivity, $q(\omega_i)$ is a frequency-dependent weighting, and the subscript ω_i denotes monochromatic data slice.

To describe $\mathbf{P}$ under a single-scattering assumption, we resort to the SRME equation (Verschuur et al., 1992) that relates $\mathbf{P}$ to $\mathbf{G}$ by multi-dimensional convolution. Following the detail-hiding "data matrix" notation (Berkhout and Pao, 1982), where multi-dimensional convolution is turned into monochromatic matrix multiplication, the total up-going wavefield can be expressed as:

$$\mathbf{P}_{\omega_i} = \mathbf{G}_{\omega_i}(\mathbf{Q}_{\omega_i} + \mathbf{R}_{\omega_i}\mathbf{P}_{\omega_i}), \tag{1}$$

where $\mathbf{R}_{\omega_i}$ is the reflectivity matrix of the surface, which is assumed to be $-\mathbf{I}$. Matrix $\mathbf{G}_{\omega_i}$ can be explicitly expressed as

$$\begin{aligned} \mathbf{G}_{\omega_i} &= \mathbf{D}_r\mathbf{H}^{-1}_{\omega_i}[\mathbf{m}](\mathbf{D}_s^*\mathbf{I}) \\ &\doteq \mathrm{vec}^{-1}(\mathbf{F}[\mathbf{m},\mathbf{I}]). \end{aligned} \tag{2}$$

Here $\mathbf{D}_r$ is a restriction operator that collects data at receiver locations, and $\mathbf{D}_s^*$ injects source wavefield at source locations. The matrix $\mathbf{H}_{\omega_i}(\mathbf{m})$ is monochromatic Helmholtz operator parameterized by medium properties described by $\mathbf{m}$. The matrix $\mathbf{I}$ represents an impulsive source array, each column of which represents one sequential source. Note that all operators are applied to $\mathbf{I}$ on a column basis so the r.h.s. of equation (2) remains a matrix. Notation "vec" means to vectorize a wavefield, and "vec^{-1}" means to undo the action of "vec". The matrix $\mathbf{F}$ denotes the modelling operator. Then by the associativity of matrix multiplication, equation (1) can be written as:

$$\begin{aligned} \mathbf{P}_{\omega_i} &= \mathrm{vec}^{-1}(\mathbf{F}[\mathbf{m},\mathbf{I}])(\mathbf{Q}_{\omega_i} - \mathbf{P}_{\omega_i}) \\ &= \mathbf{D}_r\mathbf{H}^{-1}_{\omega_i}[\mathbf{m}](\mathbf{D}_s^*\mathbf{I})(\mathbf{Q}_{\omega_i} - \mathbf{P}_{\omega_i}) \\ &= \mathbf{D}_r\mathbf{H}^{-1}_{\omega_i}[\mathbf{m}](\mathbf{D}_s^*(\mathbf{Q}_{\omega_i} - \mathbf{P}_{\omega_i})) \\ &\doteq \mathrm{vec}^{-1}(\mathbf{F}[\mathbf{m},\mathbf{Q}_{\omega_i} - \mathbf{P}_{\omega_i}]). \end{aligned} \tag{3}$$

From equation 3, we can identify $\mathbf{Q}-\mathbf{P}$ as the source term for the total up-going wavefield $\mathbf{P}$; moreover, we can directly

inject the term into wave-equation solver. This has great significance in terms of computational efficiency: first, the expensive multi-dimensional convolution is implicitly carried out by wave-equation solver; second, full source sampling required to implement multi-dimensional convolution can now be removed, which opens the opportunity to subsample over sources to reduce the number of PDE solves in wave-equation migration. In the following discussions, we denote $\mathbf{Q}-\mathbf{P}$ by $\mathbf{V}$ for simplicity.

In practice, however, until the advent of Estimate of Primary by Sparse Inversion (EPSI), recovering true amplitude source signature $\mathbf{Q}$ remained infeasible because of the amplitude ambiguity nature of blind deconvolution (van Groenestijn and Verschuur, 2009; Lin and Herrmann, 2012). EPSI can be regarded as the inversion alternative of SRME, where obtaining true amplitude $\mathbf{G}$ and $\mathbf{Q}$ are the objectives.

By ignoring internal multiples, we linearize equation (3) by the Born approximation:

$$\mathbf{p}_{\omega_i} = \nabla\mathbf{F}[\mathbf{m}_0, \mathbf{V}_{\omega_i}]\delta\mathbf{m}, \tag{4}$$

where $\mathbf{p}_{\omega_i} = \mathrm{vec}(\mathbf{P}_{\omega_i})$, $\nabla\mathbf{F}$ is the Born scattering operator, and $\mathbf{m_0}$ is a smooth background model. Then, by stacking over frequencies, we have:

$$\begin{aligned}\mathbf{p} &= \begin{bmatrix}\nabla\mathbf{F}_{\omega_1}(\mathbf{m}_0, \mathbf{V}_{\omega_i})\\ \vdots \\ \nabla\mathbf{F}_{\omega_{\mathrm{nf}}}(\mathbf{m}_0, \mathbf{V}_{\omega_i})\end{bmatrix}\delta\mathbf{m}\\ &\doteq \nabla\mathbf{F}(\mathbf{m_0}, \mathbf{v})\delta\mathbf{m},\end{aligned} \tag{5}$$

where "nf" is the number of frequencies, and $\mathbf{v} = \mathrm{vec}(\mathbf{V})$. Then, $\delta\mathbf{m}$ can be obtained by inverting $\nabla\mathbf{F}$.

FAST IMAGING WITH MULTIPLES

We start with going over the proposed method in our previous work (Tu and Herrmann, 2012). We first write down the formulation, and then discuss its elements in more details. The formulation is expressed as:

$$\begin{aligned}\delta\tilde{\mathbf{m}} &= \mathbf{C}^H \operatorname*{argmin}_{\delta\mathbf{x}} ||\delta\mathbf{x}||_1\\ &\text{subject to } ||\underline{\mathbf{p}} - \nabla\mathbf{F}[\mathbf{m_0}, \underline{\mathbf{v}}]\mathbf{C}^H\delta\mathbf{x}||_2 \le \sigma,\end{aligned} \tag{6}$$

where underlined quantities represent subsampled variables (i.e., $\underline{\mathbf{v}} = \mathbf{RMv}$, $\underline{\mathbf{p}} = \mathbf{RMp}$), $\delta\mathbf{m}$ is the vectorized model perturbation (i.e., true amplitude reflectivities). We adjust σ to allow misfit from internal multiples, incoherent noises, and modelling errors. The matrix $\mathbf{C}^H$ is 2D curvelet synthesis operator (Candès et al., 2006).

Fast inversion by subsampling

Inverting equation (5) requires solving PDEs for all sequential sources and frequencies, which is computationally very expensive if ever feasible. To reduce the number of PDE solves, we subsample over sources and frequencies (Herrmann and Li, 2011). The subsampling operator $\mathbf{RM}$ can be factorized into a mixing operator $\mathbf{M}$ that in our case blends all sequential sources into simultaneous sources, and a restriction operator $\mathbf{R}$ that uniform-randomly chooses a subset of simultaneous sources and frequencies. To be specific, the mixing operator $\mathbf{M}$ can be written as:

$$\mathbf{M} = \mathbf{I}^{\Omega} \otimes \mathbf{M}^{\Sigma} \otimes \mathbf{I}^{\Gamma}, \tag{7}$$

where $\mathbf{I}^{\Omega}$ and $\mathbf{I}^{\Gamma}$ indicate that frequencies and receivers remain untouched. Operator $\mathbf{M}^{\Sigma}$ is a square matrix with random Gaussian distributed entries. The restriction operator $\mathbf{R}$ can be written as:

$$\mathbf{R} = \mathbf{R}^{\Omega} \otimes \mathbf{R}^{\Sigma} \otimes \mathbf{I}^{\Gamma}, \tag{8}$$

implying subsampling over frequencies and sources. The frequency restriction operator $\mathbf{R}^{\Omega}$ is a flat matrix comprised of random selected rows from the identity matrix. Subsampling over frequencies is considered as a common practice in frequency domain wave-equation migration by assuming redundancies in seismic data with respect to frequencies. However, a frequency subset should still be chosen with care (Herrmann et al., 2009). The source restriction operator $\mathbf{R}^{\Sigma}$ can be similarly formulated.

Sparsity-promoting formulation

We follow the formulation by Herrmann and Li (2011) and cast the inversion as a Basis Pursuit De-Noise (BPDN) problem (Chen et al., 2001; van den Berg and Friedlander, 2008), as shown in equation (6). With this formulation, we really reap benefits from the redundancy information in multiples (Herrmann, 2010). To solve the ℓ_1 minimization problem, we use SPGℓ1 (van den Berg and Friedlander, 2008), which underlyingly solves a series of LASSO (Tibshirani, 1996; Osborne et al., 1999) sub-problems. A Newton root finding algorithm on the Pareto curve (van den Berg and Friedlander, 2008) is used to gradually relax the sparsity level, measured by τ as is in algorithm (1), until the objective residual σ is reached.

Leveraging curvelet sparsity

We incorporate curvelet transform here for two reasons. First, model perturbation $\delta\mathbf{m}$ demonstrates a strongly sparse pattern in the curvelet domain (Candès and Donoho, 2004), which complements our sparsity-promoting formulation (Herrmann et al., 2008; Herrmann and Li, 2011). Second, curvelet atoms decay smoothly in the physical domain, which is crucial to retain continuity in reconstructing the reflectivities by sparse inversion. Similar observations in wavefield inversion are demonstrated by Lin and Herrmann (2012).

Synthetic experiment settings

We use a synthetic salt dome model to verify the proposed method. The dimensions of the model are 1250m vertically and 2235m horizontally with 5m grid distance. We make linearized data according to equation (5) assuming known source wavefield $\mathbf{v}$. The total recording time is 2.044s with 4ms sampling interval, which corresponds to a total number of 122 nonzero frequencies in 0-60Hz frequency range. There are 150 co-located sources and receivers with 15m spacing. The result using all sequential sources and frequencies is shown in figure (1), which will be used as a reference for the following experiments. To demonstrate the effectiveness of the proposed method, the result using 10 simultaneous sources and all frequencies is shown in figure (2). Both images are obtained after 60 SPGℓ1 iterations. By using simultaneous sources, we

achieve 15X speed-up without compromising on image quality. Throughout this abstract, computational cost is measured by number of PDE solves.

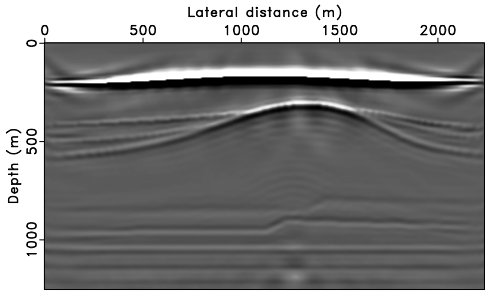

Figure 1: Result using all sources and frequencies

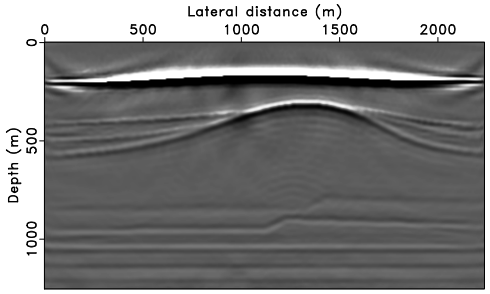

Figure 2: Result using 10 simultaneous sources and all frequencies

BREAKING THE CORRELATIONS

As we continue to decrease the number of sources and frequencies, more migration artifacts occur. An example is shown in figure (3), where the image is overwhelmed by noisy artifacts. This result is obtained after 305 SPGℓ1 iterations using only two simultaneous sources and 15 frequencies. A rough estimate of the total number of PDE solvers is 36600 (*rough* in the sense that SPGℓ1 sometimes does more than two matrix-vector multiplications — one forward and one adjoint — per iteration), which is equal to the number of PDE solves of a single reverse-time migration using all sources and all frequencies. In the following experiments, we use the same number of simultaneous sources, frequencies, and iterations.

To mitigate these subsampling artifacts, we borrow insights from recent advances in approximate message passing (AMP) (Donoho et al., 2009), and draw a new subsampling operator each time we finish solving a LASSO sub-problem (Herrmann, 2012). The computational steps are described in algorithm 1.

Renewal over sources

Renewal over sources have been discussed by Herrmann and Li (2011); Tu and Herrmann (2012). This case can be regarded as redrawing the mixing operator $\mathbf{M}$, which breaks the correlation built up between model iterate and the encoding Gaussian matrix $\mathbf{M}^{\Sigma}$ in equation (7). The resulting image is shown in figure (4). We can see that the artifacts in figure (3) are largely removed, without extra computational expense.

Algorithm 1 Fast imaging by subsampling and renewal

Input:
target residual σ
seismic data in frequency domain $\mathbf{P}$
the corresponding source in frequency domain $\mathbf{V}$
initialization:
$k \leftarrow 0$, initial guess $\delta\mathbf{x}_k \leftarrow$ zero vector
repeat
  $\underline{\mathbf{p}}_{k+1}$ and $\underline{\mathbf{v}}_{k+1} \leftarrow$ redrawing subsampling operator
  $\tau_k \leftarrow ||\delta\mathbf{x}_k||_1$
  $\tau_{k+1} \leftarrow$ determine from $\tilde{\sigma}$ and τ_k using Newton's method on the Pareto curve
  $\delta\mathbf{x}_{k+1} \leftarrow \text{argmin}_{\delta\mathbf{x}} ||\underline{\mathbf{p}} - \nabla\mathbf{F}[\mathbf{m_0}, \underline{\mathbf{v}}]\mathbf{C}^H\delta\mathbf{x}||_2$
   s. t. $||\delta\mathbf{x}||_1 \leq \tau_{k+1}$ //warm start with $\delta\mathbf{x}_k$
  $k \leftarrow k+1$, $\delta\mathbf{x}_k \leftarrow \delta\mathbf{x}_{k+1}$
until $||\underline{\mathbf{p}} - \nabla\mathbf{F}(\mathbf{m_0}, \underline{\mathbf{v}})\mathbf{C}^H\delta\mathbf{x}||_2 \leq \sigma$
or a predetermined iteration limit is reached
Output: estimated model perturbation $\delta\mathbf{m} = \mathbf{C}^H\delta\mathbf{x}$

Renewal over frequencies

One difference between source and frequency subsampling is that for simultaneous sources, the illumination power of each sequential source is more or less constant, but for frequencies, we lose the frequency components that are not sampled. Therefore, we propose to redraw the frequency restriction operator $\mathbf{R}^{\Omega}$ also. As a result, we potentially make use of all frequency contents in the data. The resulting image is shown in figure (5). We also observe significant improvements over figure (3).

Renewal over both sources and frequencies

To maximize the benefits of renewal, we propose to renew over sources and frequencies simultaneously . The resulting image is shown in figure (6). We can see that the residual artifacts observed in figure (4) and (5) are further reduced. Compared with the reference result (figure (1)), we achieve 120X speed up.

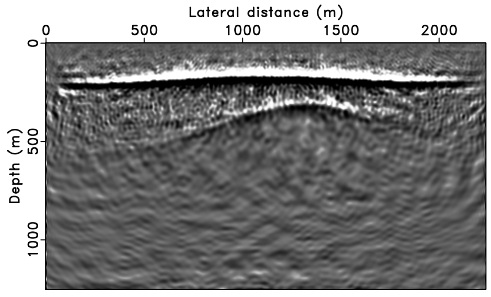

Figure 3: Result using 2 simultaneous sources and 15 frequencies without renewal

Insights from the solution path

The solution paths, represented as the relative ℓ_2-norm residual in equation (6) being a function of the ℓ_1-norm of the so-

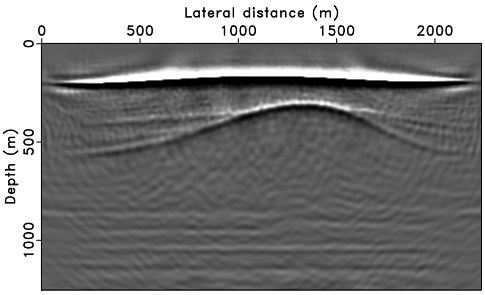

Figure 4: Result using 2 simultaneous sources and 15 frequencies with source renewal

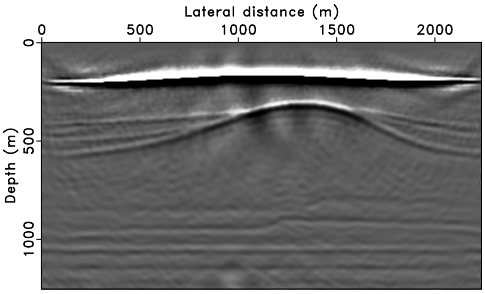

Figure 5: Result using 2 simultaneous sources and 15 frequencies with frequency renewal

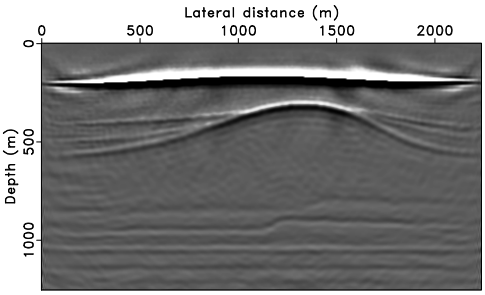

Figure 6: Result using 2 simultaneous sources and 15 frequencies with both renewal

lution $\delta\mathbf{x}$, are drawn in figure (7). The relative model errors, measured by the ℓ_2-norm of the difference between model updates and the reference model (figure (1)), are drawn in figure (8). The outliers in figure (8) are intermediate estimates during line-searches.

From the solution path plots, we can see that despite all the strong artifacts in figure (3), the scenario without any renewal surprisingly obtains the fastest residual decrease, all the way down to almost zero. However, the model error plots reveal that smaller residual does not necessarily mean smaller model error, which implies that the artifacts in figure (3) live in the null space of the the initially subsampled Born operator. Therefore these artifacts cannot be removed by simply adding more

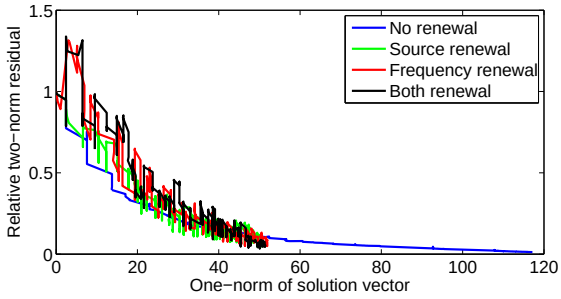

Figure 7: Solution paths of the four scenarios

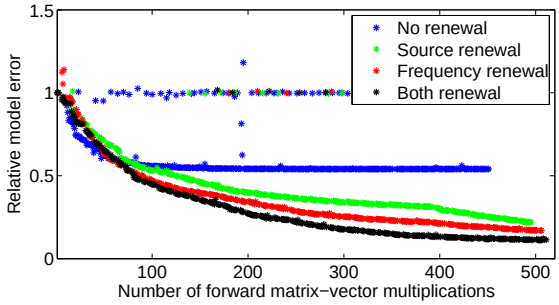

Figure 8: Model errors of the four scenarios

iterations. We can see from figure (8) that the model error stops decreasing at some point in the no-renewal scenario, but continues to decrease in any other scenario with renewal.

We conclude that by changing simultaneous sources and frequency subsets, the null space of the subsampled Born operator also changes. Then the artifacts that no longer live in the null space of the newly subsampled Born operator can be removed. From this point of view, we really benefit from renewals by being able to use a *significantly smaller* batch size combined with *moderately more* iterations. As a result, the computation is still less expensive than it would otherwise be. Furthermore, using a smaller batch size also saves computer memory.

CONCLUSION

In this abstract, we demonstrate how renewal over sources and frequencies can remove the artifacts arising from insufficient sampling, which in turn enables us to further decrease the subsampling ratio towards faster inversions.

ACKNOWLEDGEMENTS

We thank Eric Verschuur for providing the salt dome model. Ning Tu is partially sponsored by the Chinese Scholarship Council. This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.

REFERENCES

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.

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: Multiscale Modeling and Simulation, **5**, no. 3, 861–899.

Candès, E. J., and D. L. Donoho, 2004, New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities: Communications on Pure and Applied Mathematics, **57**, 219–266.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: SIAM Review, **43**, 129.

Donoho, D., 2006, Compressed sensing: Information Theory, IEEE Transactions on, **52**, 1289 –1306.

Donoho, D., A. Maleki, and A. Montanari, 2009, Message passing algorithms for compressed sensing: Proceedings of the National Academy of Sciences, **106**, 18914–18919.

Herrmann, F. J., 2010, Randomized sampling and sparsity: Getting more information from fewer samples: Geophysics, **75**, WB173–WB187.

———, 2012, Pass on the message: recent insights in large-scale sparse recovery: Presented at the EAGE technical program, EAGE, EAGE.

Herrmann, F. J., Y. A. Erlangga, and T. T. Lin, 2009, Compressive simultaneous full-waveform simulation: Geophysics, **74**, A35.

Herrmann, F. J., and X. Li, 2011, Efficient least-squares imaging with sparsity promotion and compressive sensing: Tech. Rep. TR-2011-03, Department of Earth and Ocean Sciences, University of British Columbia, Vancouver.

Herrmann, F. J., D. Wang, G. Hennenfent, and P. P. Moghaddam, 2008, Curvelet-based seismic data processing: a multiscale and nonlinear approach: Geophysics, **73**, A1–A5.

Lin, T. T., and F. J. Herrmann, 2012, Robust estimation of primaries by sparse inversion via one-norm minimization. submitted to Geophysics, March 12, 2012.

Lin, T. T., N. Tu, and F. J. Herrmann, 2010, Sparsity-promoting migration from surface-related multiples: SEG expanded abstracts, 3333–3337.

Liu, Y., X. Chang, D. Jin, R. He, H. Sun, and Y. Zheng, 2011, Reverse time migration of multiples for subsalt imaging: , WB209–WB216.

Osborne, M. R., B. Presnell, and B. A. Turlach, 1999, On the lasso and its dual: Journal of Computational and Graphical Statistics, **9**, 319–337.

Tibshirani, R., 1996, Regression shrinkage and selection via the lasso: Journal of the Royal Statistical Society Series B Methodological, **58**, 267–288.

Tu, N., and F. J. Herrmann, 2012, Least-squares migration of full wavefield with source encoding: Presented at the EAGE technical program, EAGE.

Tu, N., T. T. Lin, and F. J. Herrmann, 2011, Sparsity-promoting migration with surface-related multiples: Presented at the EAGE technical program, EAGE.

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, Estimating primaries by sparse inversion and application to near-offset data reconstruction: Geophysics, **74**, A23–A28.

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

———, 2011, Seismic migration of blended shot records with surface-related multiple scattering: Geophysics, **76**, A7–A13.

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177.

Whitmore, N. D., A. A. Valenciano, and W. Sollner, 2010, Imaging of primaries and multiples using a dual-sensor towed streamer: SEG Expanded Abstracts, 3187–3192.