

Source separation via SVD-free rank minimization in the hierarchical semi-separable representation

Haneet Wason*, Rajiv Kumar*, Aleksandr Y. Aravkin[†] and Felix J. Herrmann*

*Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia, [†]IBM T.J. Watson Research Center, New York, USA

SUMMARY

Recent developments in matrix rank optimization have allowed for new computational approaches in the field of source separation or deblending. In this paper, we propose a source separation algorithm for blended marine acquisition, where two sources are deployed at different depths (over/under acquisition). The separation method incorporates the Hierarchical Semi-Separable structure (HSS) inside rank-regularized least-squares formulations. The proposed approach is suitable for large scale problems, since it avoids SVD computations and uses a low-rank factorized formulation instead. We illustrate the performance of the new HSS-based deblending approach by simulating an over/under blended acquisition, wherein uniformly random time delays (of < 1 second) are applied to the one of the sources.

INTRODUCTION

The benefits of acquiring and processing over/under data are clear; the recorded bandwidth is extended at both low and high ends of the spectrum. The over/under acquisition method allows separation of the up- and downgoing wavefields at the source (or receiver) using a vertical pair of sources (or receivers) to determine wave direction (Moldoveanu et al., 2007). Acquiring data with over/under sources is an instance of simultaneous acquisition, which generates blended datasets that need to be separated into the individual, unblended datasets for further processing.

The challenge of source separation or deblending has been addressed by many researchers (Stefani et al., 2007; Moore et al., 2008; Akerberg et al., 2008; Huo et al., 2009), wherein the key observation has been that as long as the sources are fired at suitably randomly dithered times, the resulting interferences (or source crosstalk) will appear noise-like in specific gather domains such as common-offset and common-receiver, turning the separation problem into a (random) noise removal procedure. Inversion-type algorithms (Moore, 2010; Abma et al., 2010; Mahdad et al., 2011; Doulgeris et al., 2012; Baardman and van Borselen, 2013) take advantage of sparse representations of coherent seismic signals. Wason and Herrmann (2013a,b) proposed an alternate sampling strategy for simultaneous acquisition (*time-jittered marine*) that leverages ideas from compressed sensing (CS), addressing the deblending problem through a combination of tailored (blended) acquisition design and sparsity-promoting recovery via convex optimization using ℓ_1 constraints.

More recently, rank-minimization based techniques have been used for source separation by Maraschini et al. (2012) and Cheng and Sacchi (2013). The general idea is to exploit the low-rank structure of seismic data when it is organized in a ma-

trix. Low-rank structure refers to the small number of nonzero singular values, or quickly decaying singular values. Maraschini et al. (2012) followed the rank-minimization approach proposed by Oropenza and Sacchi (2011), who identified that seismic temporal frequency slices organized into a block Hankel matrix, in ideal conditions, is a matrix of rank k , where k is the number of different plane waves in the window of analysis. These authors showed that additive random noise increase the rank of the block Hankel matrix, and the authors presented an iterative algorithm that resembles seismic data reconstruction with the method of projection onto convex sets, where they use a low-rank approximation of the Hankel matrix via the randomized singular value decomposition (Liberty et al. (2007); Halko et al. (2011a)) to interpolate seismic temporal frequency slices. While this technique may be effective interpolating data with a limited number of distinct dips, the approach requires embedding the data into an even larger space where each dimension of size n is mapped to a matrix of size $n \times n$ and that forces them to work on small windows and now choosing the size of these windows becomes a factor.

In this paper, we avoid the direct approach to rank-minimization that involves computing singular value decompositions (SVD) of the matrices and follow instead the approach proposed by Aravkin et al. (2013); Kumar et al. (2013) and the references therein. The key idea here is that the monochromatic frequency slices of the fully sampled data matrix have low-rank structure in the midpoint-offset (m-h) domain, whereas, data with random noise increases the rank of the resulting frequency slice in the m-h domain. Seismic frequency slices exhibit low-rank structure in the m-h domain at low-frequencies, but not at high-frequencies. This behaviour is due to the increase in oscillations as we move from low to high-frequency slices in the m-h domain, even though the energy remains focused around the diagonal (Kumar et al., 2013).

Rank-minimization in the high-frequency range requires extended formulations that incorporate low-rank structure. While Engquist and Ying (2010) propose to overcome this difficulty by proposing directional low-rank approximations for Green's functions of the acoustic wave equation in 3D, we rely instead on the Hierarchically Semi-Separable matrix representation (HSS) method proposed by Chandrasekaran et al. (2006) to represent frequency slices. The key idea in the HSS representation is that certain full-rank matrices, e.g., matrices that are diagonally dominant with energy decaying along the off-diagonals, can be represented by a collection of low-rank submatrices.

Following the same ideas, in an over/under simultaneous acquisition, we observe that monochromatic frequency slices of the fully sampled blended data matrix, with periodic firing times, have low-rank structure in the m-h domain, whereas, uniformly random time delays applied to one or both sources,

increase the rank of the resulting frequency slice in the m-h domain. Hence, the blended data in over/under randomized acquisition is separated into its constituent source components using a fast optimization approach that combines the (SVD-free) matrix factorization approach recently developed by Lee et al. (2010) with the Pareto curve approach proposed by Berg and Friedlander (2008).

The paper is organized as follows. We start by explaining the methodology to perform source separation by combining the HSS structure with factorization-based rank-regularized optimization formulations (Lee et al., 2010). Next, we demonstrate the successful implementation of the proposed source separation strategy on a simulated dataset. In addition, we also make comparisons with the sparsity-promoting deblending techniques.

THEORY

The success of rank-minimization hinges on the fact that regularly sampled target dataset should exhibit a low-rank structure in rank-revealing "transform domain". Kumar et al. (2013) showed that the frequency slices of unblended seismic data do not exhibit a low-rank structure in the source-receiver (s-r) domain since strong wavefronts extend diagonally across the s-r plane. However, transforming the data into the m-h domain results in a vertical alignment of the wavefronts, thereby reducing the rank of the frequency slice matrix. In the low frequency slices, this vertical alignment can be accurately approximated by a low-rank representation. On the other hand, high frequency slices include a variety of wave oscillations that increase the rank. However, Kumar et al. (2013) mentioned that it is possible to find accurate low-rank approximations of submatrices of the high-frequency slices by partitioning the data into a Hierarchical Semi-Separable (HSS) structure. The HSS structure first partitions a matrix into diagonal and off-diagonal submatrices. The same partitioning structure is then applied recursively to the diagonal sub-matrices only, since these sub-matrices still possess high rank structure.

In this paper, we consider an over/under blended acquisition scenario with two sources placed at different depths with an interval of 5 m. We observe that the monochromatic frequency slices of the blended data, with periodic firing times of sources, have low-rank structure in the m-h domain, while the frequency slices of the blended data with uniformly random firing time delays applied to the second source, do not. Note, that we have fully sampled data in both cases. Figures 1 (a,b) show the monochromatic frequency slice at 25 Hz in the m-h domain of the blended data with periodic firing times of both sources, and with uniformly random firing time delays of the second source, respectively.

Randomization of the time delays increases the rank or makes the singular values decay less quickly in the m-h domain, an essential feature for rank-minimization techniques to be effective. To illustrate this behaviour, we plot the decay of the singular values of blended data in Figure 2. Note that uniformly random firing time delays does not noticeably change the decay of the singular values in the s-r domain, as expected,

but significantly slow down the decay rate in the m-h domain as shown in Figures 2(a) and 2(b). Therefore, low-frequency slices of unblended datasets can be reconstructed by solving the rank-minimization problem in the m-h domain, while high-frequency slices can be reconstructed by first performing an HSS partitioning and then solving the rank-minimization problem on each partition in its respective m-h domain (Kumar et al., 2013).

Source separation

Let X_0 be a matrix in $\mathbb{C}^{n \times m}$ and let \mathcal{A} be a linear measurement operator that maps from $\mathbb{C}^{n \times m} \rightarrow \mathbb{C}^p$ with $p \ll n \times m$. Recht et al. (2010) showed that under certain general conditions on the operator \mathcal{A} , the solution to the rank-minimization problem can be found by solving the following nuclear-norm minimization problem:

$$\min_X \|X\|_* \quad \text{s.t. } \|\mathcal{A}(X) - b\|_2 \leq \varepsilon, \quad (\text{BPDN}_\varepsilon)$$

where b is a set of measurements, $\|X\|_* = \|\sigma\|_1$, and σ is the vector of singular values. In the case of over/under simultaneous acquisition, given two unblended data matrices X_1 and X_2 and blended data measurements b , we can redefine our systems of equations as

$$\underbrace{\begin{bmatrix} M\mathcal{S}^H & MT\mathcal{S}^H \end{bmatrix}}_{\mathcal{A}} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = b, \quad (1)$$

where M is the sampling operator, \mathcal{S} is the transformation operator from the s-r domain to the m-h domain, H denotes the Hermitian transpose, and T is defined as a time delay operator which applies the uniformly random time delays to the second source.

In order to efficiently solve $(\text{BPDN}_\varepsilon)$, we use an extension of the SPGL_1 solver (Berg and Friedlander, 2008) developed for the $(\text{BPDN}_\varepsilon)$ problem in Aravkin et al. (2012). The SPGL_1 algorithm finds the solution to the $(\text{BPDN}_\varepsilon)$ by solving a sequence of LASSO subproblems

$$\min_X \|\mathcal{A}(X) - b\|_2 \quad \text{s.t. } \|X\|_* \leq \tau, \quad (\text{LASSO}_\tau)$$

where τ is updated by traversing the Pareto curve. Solving each LASSO subproblem requires a projection onto the nuclear norm ball $\|X\|_* \leq \tau$ in every iteration by performing a singular value decomposition and then thresholding the singular values. In the case of large scale seismic problems, it becomes prohibitively expensive to carry out such a large number of SVDs. Instead, we adopt a recent factorization-based approach to nuclear-norm minimization (Rennie and Srebro, 2005; Lee et al., 2010; Recht and Ré, 2011). The factorization approach parametrizes the matrix $(X_1, X_2) \in \mathbb{C}^{n \times m}$ as the product of two low rank factors $(L_1, L_2) \in \mathbb{C}^{n \times k}$ and $(R_1, R_2) \in \mathbb{C}^{m \times k}$, such that,

$$X = \begin{bmatrix} L_1 R_1^H \\ L_2 R_2^H \end{bmatrix} \quad (2)$$

The optimization scheme can then be carried out using the factors (L_1, L_2) and (R_1, R_2) instead of (X_1, X_2) , thereby significantly reducing the size of the decision variable from $2nm$ to

$2k(n+m)$ when $k \ll m, n$. Rennie and Srebro (2005) showed that the nuclear norm obeys the relationship

$$\|X\|_* \leq \frac{1}{2} \left\| \begin{bmatrix} L_1 \\ R_1 \end{bmatrix} \right\|_F^2 + \frac{1}{2} \left\| \begin{bmatrix} L_2 \\ R_2 \end{bmatrix} \right\|_F^2 =: \Phi(L_1, R_1, L_2, R_2), \quad (3)$$

where $\|\cdot\|_F$ is the Frobenius norm of the matrix (sum of the squared entries).

Consequently, the LASSO subproblem can be replaced by

$$\min_{L_1, R_1, L_2, R_2} \|\mathcal{A}(X) - b\|_2 \quad \text{s.t.} \quad \Phi(L_1, R_1, L_2, R_2) \leq \tau, \quad (4)$$

where the projection onto $\Phi(L_1, R_1, L_2, R_2) \leq \tau$ is easily achieved by multiplying each factor (L_1, L_2) and (R_1, R_2) by the scalar $\sqrt{2\tau/\Phi(L, R)}$. By equation (3) for each HSS sub-matrix in the m-h domain, we are guaranteed that $\|X\|_* \leq \tau$ for any solution of (4). Once the optimization problem is solved, we transform each sub-matrix back from the m-h to the s-r domain, where we concatenate all the sub-matrices to get the deblended monochromatic frequency data matrices. One of the advantages of HSS representation is that it works with the recursive partitioning of a matrix, and sub-matrices can be solved in parallel, speeding up the optimization formulation.

NUMERICAL RESULTS

We illustrate the performance of our proposed algorithm on data generated from the BG compass model using the IWAVE software. With a source (and) receiver sampling of 25.0 m, one dataset is generated with a source-depth of 10.0 m (Figure 3a), while the other dataset has the source at 15.0 m depth. Each dataset has $N_t = 1024$ time samples, $N_r = 129$ receivers and $N_s = 129$ shots. To simulate the over/under blended acquisition scenario, the two datasets are (simply) summed for the periodic case, while uniformly random time delays between 0-1 second are applied to the dataset of the second source (Figure 3b) for the randomized case. Figure 3c shows the resulting blended shot gather. Since the time delays are less than a second, we call this scenario the *low variability* randomized acquisition. Using the proposed rank-minimization technique, we are able to successfully deblend the blended shot gathers into its constituent source components as illustrated in Figures 4(a,b). Figures 4(e,f) show the difference between the unblended shot gathers and the (recovered) deblended shot gathers of source 2 and source 1, respectively.

It is well known that seismic data admit sparse representations by curvelets that capture “wavefront sets” efficiently (Smith, 1998; Candès and Demanet, 2005; Candès et al., 2006; Herrmann et al., 2008). Therefore, we also compare the proposed rank-minimization based deblending algorithm with the sparsity-promoting deblending technique. Here, we use (BPDN $_{\epsilon}$) formulation to minimize the ℓ_1 norm instead of the nuclear norm, and the transformation operator \mathcal{S} is the 2-D curvelet operator. Figures 4(c,b) show the deblended shot gathers using sparsity-promoting based techniques and Figures 4(g,h) show the corresponding difference plots between the unblended shot gathers and the (recovered) deblended shot gathers of source 2 and source 1, respectively. It can be clearly seen that source separation via sparsity-promotion does not help to deblend the

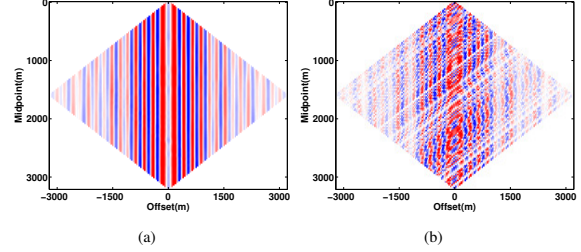


Figure 1: Monochromatic frequency slice at 25 Hz in the m-h domain for blended data (a) with periodic firing times for both sources, and (b) with uniformly random firing times of the second source.

blended dataset especially in this low variability scenario. To understand this behaviour, we plot the decay of the curvelet coefficients (Figure 2c) of the monochromatic frequency slice at 25 Hz in the s-r domain, for the periodic and the randomized blended acquisition. For this low variability scenario, the randomization does not destroy the sparse structure of the data, i.e., does not decrease the decay of the curvelet coefficients significantly, which is an essential requirement for sparsity-promoting deblending techniques as shown in Wason and Herrmann (2013a,b), which can be termed as the *high variability* randomized acquisition scenario.

CONCLUSIONS

We have presented a new method for source separation in an over/under acquisition scenario, where we incorporate the HSS structure with factorization-based rank-regularized optimization formulations. We combine the Pareto curve approach for optimizing (BPDN $_{\epsilon}$) formulations with the SVD-free matrix factorization methods to solve the nuclear-norm optimization formulation. The proposed rank-minimization source separation technique helps in deblending the blended datasets in the low variability randomized acquisition scenarios as observed here. The experimental results demonstrate the potential benefit of this methodology. Incorporating HSS into this scheme is very promising, since we can attack high-rank structures using the partitioning, and in addition gain a computational advantage by optimizing each sub-block in parallel. In reality, seismic data are typically irregularly sampled along spatial axes, therefore, future work includes working with non-uniform sampling grids for 2-D and 3-D seismic.

ACKNOWLEDGMENTS

We would like to thank the authors of IWAVE (a framework for wave simulation). This work was financially supported in part by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN 261641-06) and the Collaborative Research and Development Grant DNOISE II (CDRP J 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, CGG, ION GXT, Petrobras, PGS, Statoil, Total SA, WesternGeco, Woodside.

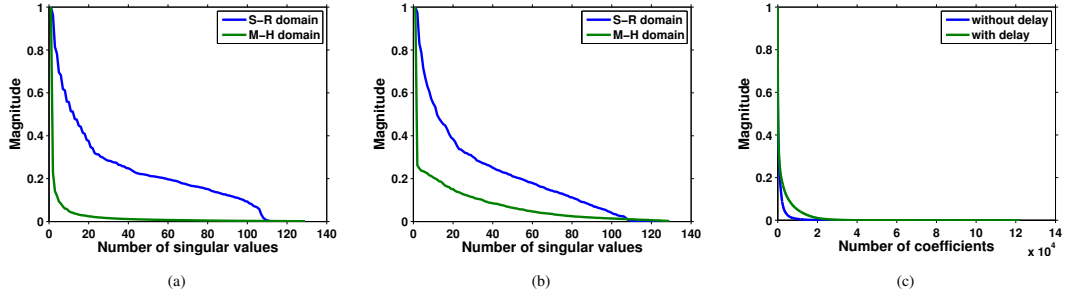


Figure 2: Decay of singular values for a frequency slice (at 25 Hz) of blended shot gather, with (a) periodic, and (b) uniformly random firing time delays of the second source. (c) Decay of the curvelet coefficients in s-r domain for case (a,b).

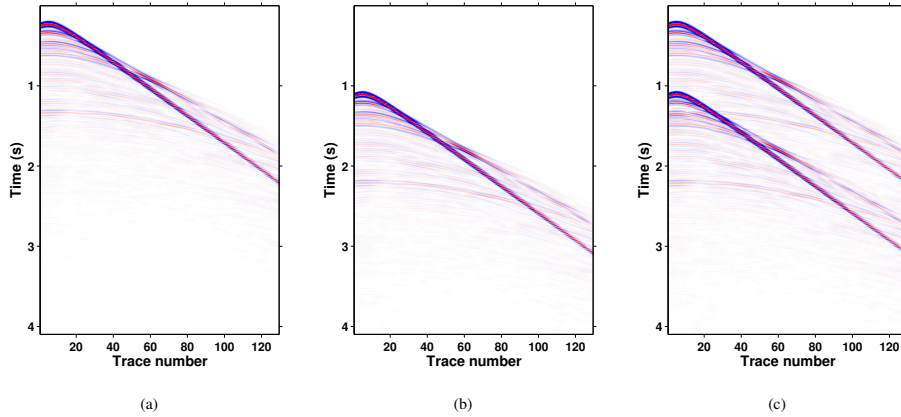


Figure 3: (a) Unblended shot gather of source 1, (b) Unblended, uniformly random time-delayed shot gather of source 2, (c) Blended shot gather.

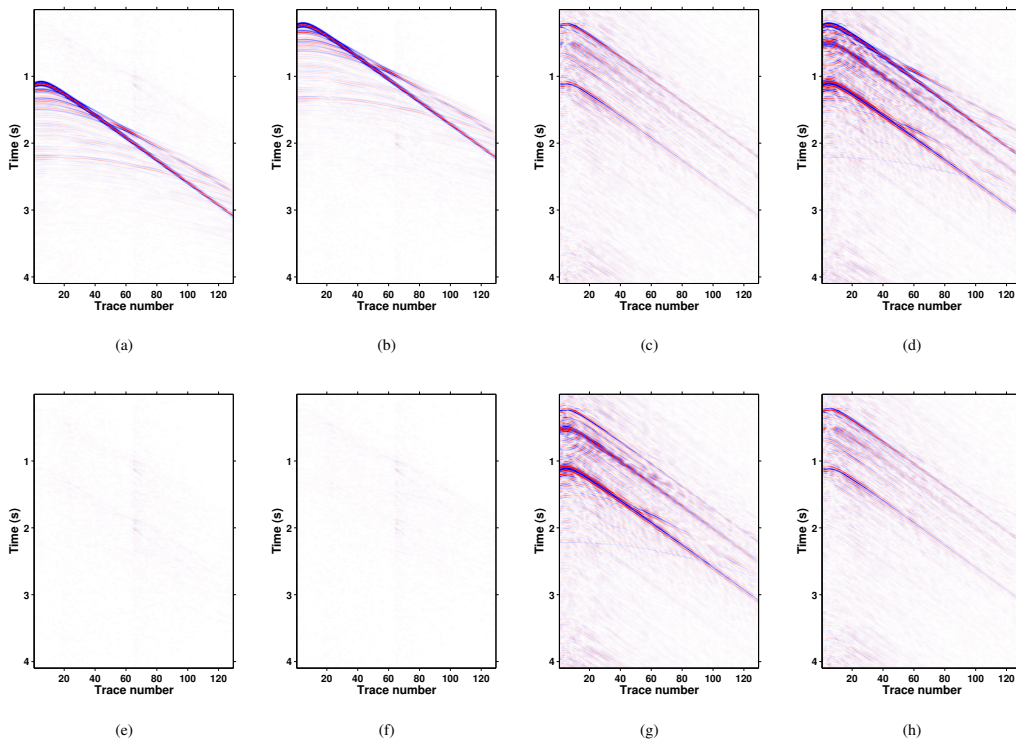


Figure 4: Deblended shot gathers of source 2 and source 1 using HSS based rank-minimization (a,b) and using curvelet based sparsity-promoting technique (c,d). (e,f,g,h) Difference of deblended shot gathers from unblended shot gathers (Figure 3 (a,b)).

REFERENCES

- Abma, R., T. Manning, M. Tanis, J. Yu, and M. Foster, 2010, High quality separation of simultaneous sources by sparse inversion: Presented at the 72nd EAGE Conference and Exhibition.
- Akerberg, P., G. Hampson, J. Rickett, H. Martin, and J. Cole, 2008, Simultaneous source separation by sparse radon transform: SEG Technical Program Expanded Abstracts, **27**, 2801–2805.
- Aravkin, A., J. Burke, and M. Friedlander, 2012, Variational properties of value functions: submitted to SIAM J. Opt., arXiv:1211.3724.
- Aravkin, A. Y., R. Kumar, H. Mansour, B. Recht, and F. J. Herrmann, 2013, Fast methods for denoising matrix completion formulations, with application to robust seismic data interpolation.
- Baardman, R. H., and R. G. van Borselen, 2013, Method and system for separating seismic sources in marine simultaneous shooting acquisition: Patent Application, **EP 2592439 A2**.
- Berg, E. v., and M. P. Friedlander, 2008, Probing the pareto frontier for basis pursuit solutions: SIAM Journal on Scientific Computing, **31**, 890–912.
- Candès, E., L. Demanet, D. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: Multiscale Modeling & Simulation, **5**, 861–899.
- Candès, E. J., and L. Demanet, 2005, The curvelet representation of wave propagators is optimally sparse: Communications on Pure and Applied Mathematics, **58**, 1472–1528.
- Chandrasekaran, S., P. Dewilde, M. Gu, W. Lyons, and T. Pals, 2006, A fast solver for hss representations via sparse matrices: SIAM J. Matrix Analysis Applications, **29**, 67–81.
- Cheng, J., and M. D. Sacchi, 2013, Separation of simultaneous source data via iterative rank reduction: SEG Technical Program Expanded Abstracts, 88–93.
- Doulgeris, P., K. Bube, G. Hampson, and G. Blacquiere, 2012, Convergence analysis of a coherency-constrained inversion for the separation of blended data: Geophysical Prospecting, **60**, 769–781.
- Engquist, B., and L. Ying, 2010, Fast directional algorithms for the helmholtz kernel: J. Comput. Appl. Math., **234**, 1851–1859.
- Herrmann, F. J., P. P. Moghaddam, and C. C. Stolk, 2008, Sparsity- and continuity-promoting seismic imaging with curvelet frames: Journal of Applied and Computational Harmonic Analysis, **24**, 150–173. (doi:10.1016/j.acha.2007.06.007).
- Huo, S., Y. Luo, and P. Kelamis, 2009, Simultaneous sources separation via multi-directional vector-median filter: SEG Technical Program Expanded Abstracts, **28**, 31–35.
- Kumar, R., H. Mansour, A. Y. Aravkin, and F. J. Herrmann, 2013, Reconstruction of seismic wavefields via low-rank matrix factorization in the hierarchical-separable matrix representation: Presented at the SEG.
- Lee, J., B. Recht, R. Salakhutdinov, N. Srebro, and J. Tropp, 2010, Practical large-scale optimization for max-norm regularization: Presented at the Advances in Neural Information Processing Systems, 2010.
- Mahdad, A., P. Doulgeris, and G. Blacquiere, 2011, Separation of blended data by iterative estimation and subtraction of blending interference noise: Geophysics, **76**, Q9–Q17.
- Maraschini, M., R. Dyer, K. Stevens, and D. Bird, 2012, Source separation by iterative rank reduction - theory and applications: Presented at the 74th EAGE Conference and Exhibition.
- Moldoveanu, N., L. Combee, M. Egan, G. Hampson, L. Sydora, and W. Abriel, 2007, Over/under towed-streamer acquisition: A method to extend seismic bandwidth to both higher and lower frequencies: The Leading Edge, **26**, 41–58.
- Moore, I., 2010, Simultaneous sources - processing and applications: Presented at the 72nd EAGE Conference and Exhibition.
- Moore, I., B. Dragoset, T. Ommundsen, D. Wilson, C. Ward, and D. Eke, 2008, Simultaneous source separation using dithered sources: SEG Technical Program Expanded Abstracts, **27**, 2806–2810.
- Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, V25–V32.
- Recht, B., M. Fazel, and P. Parrilo, 2010, Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization.: SIAM Review, **52**, 471–501.
- Recht, B., and C. Ré, 2011, Parallel stochastic gradient algorithms for large-scale matrix completion: Presented at the Optimization Online.
- Rennie, J. D. M., and N. Srebro, 2005, Fast maximum margin matrix factorization for collaborative prediction: Proceedings of the 22nd international conference on Machine learning, ACM, 713–719.
- Smith, H. F., 1998, A hardy space for fourier integral operators: Journal of Geometric Analysis, **8**, 629–653.
- Stefani, J., G. Hampson, and F. Herkenhoff, 2007, Acquisition using simultaneous sources: Presented at the 69th EAGE Conference and Exhibition.
- Wason, H., and F. J. Herrmann, 2013a, Ocean bottom seismic acquisition via jittered sampling: Presented at the EAGE.
- , 2013b, Time-jittered ocean bottom seismic acquisition: Presented at the SEG.