Dimensionality-reduced estimation of primaries by sparse inversion

Bander Jumah and Felix J. Herrmann, University of British Columbia, Canada

SUMMARY

Data-driven methods—such as the estimation of primaries by sparse inversion—suffer from the 'curse of dimensionality', which leads to disproportional growth in computational and storage demands when moving to realistic 3-D field data. To remove this fundamental impediment, we propose a dimensionality reduction technique where the 'data matrix' is approximated adaptively by a randomized low-rank approximation. Compared to conventional methods, our approach has the advantage that the cost of the low-rank approximation is reduced significantly, which may lead to considerable reductions in storage and computational costs of the sparse inversion. Application of the proposed formalism to synthetic data shows that significant improvements are achievable at low computational overhead required to compute the low-rank approximations.

THEORY

The success of estimating primaries by sparse inversion (EPSI, van Groenestijn and Verschuur, 2009; Lin and Herrmann, 2009; Lin et al., 2010; Lin and Herrmann, 2011)—a particular instance sparsity-promoting wavefield inversion (Herrmann and Wang, 2008)—hinges on three key components, namely (i) the existence of a fast sparsifying transform that explores structure of the surface-free Green's function, e.g. by the discrete curvelet transform; (ii) a large-scale solver that promotes sparsity, e.g. spectral-projected gradients (SPGℓ_1 Berg and Friedlander, 2008); and (iii) fast evaluation of the monochromatic data matrix and its adjoint. Of these three components, the evaluation of the action of the data matrix is by far the most challenging because it is expensive—the data matrix is full—and requires access to all data. To make things worse, this data matrix is applied iteratively during the inversion. In this paper, we propose a new approach to address this issue by using a low-rank approximation of the data matrix. Before going into detail, let us first briefly review our sparsity-promoting formulation for the estimation of primaries.

Sparsity-promoting wavefield inversion

To exploit the multi-dimensional geometry exhibited by seismic wavefields, we take Berkhout's monochromatic matrix notation—where monochromatic shot gathers are collected in the columns of a matrix—and cast this formulation into a matrix-vector form that is amenable to a solution by a sparsity-promoting program. For this purpose, consider the following linear relationship,

$$\widehat{\mathbf{G}}_i \widehat{\mathbf{U}}_i \approx \widehat{\mathbf{V}}_i, i = 1 \cdots n_f, \tag{1}$$

between the two known discretized monochromatic wavefields $\widehat{\mathbf{U}}_i$ and $\widehat{\mathbf{V}}_i$, and the unknown wavefield $\widehat{\mathbf{G}}_i$ at angular frequency $\omega = (i-1)\Delta\omega$, $i = 1 \cdots n_f$ with $\Delta\omega$ the sampling rate in the Fourier domain and n_f the number of frequencies. For simplicity, we assume the source and receiver coordinates to be co-located so that all matrices in Equation 1 are square. When possible we will also drop the frequency subscript in our notation. Finally, we used the $\approx$ symbol to indicate the presence of noise.

Even though the monochromatic data matrices are square, they are not full rank because data has a finite aperture. In addition, the data and hence the data matrix is scaled by the source wavelet, which scales frequencies at the low and high-ends of the spectrum. As a consequence, data matrices $\widehat{\mathbf{U}}_i$, $i = 1 \cdots n_f$ are ill conditioned and challenging to invert because of instabilities related to small singular values.

A common practice to counter these instabilities is to impose an energy penalty on the solution through damped least-squares. This regularization has the advantage that it yields an explicit expression for the inverse (denoted by the $\widetilde{}$ symbol) reading

$$\widetilde{\widehat{\mathbf{G}}}_i \approx \widehat{\mathbf{V}}_i \widehat{\mathbf{U}}_i^* \left(\widehat{\mathbf{U}}_i \widehat{\mathbf{U}}_i^* + \varepsilon_i^2 \mathbf{I} \right)^{-1}, i = 1 \cdots n_f, \tag{2}$$

with the symbol * denoting the conjugate transpose and ε_i a frequency-dependent regularization parameter, which controls the data misfit versus the energy penalty on $\widehat{\mathbf{G}}$.

Unfortunately, this monochromatic formulation has a number of problems. First, minimizing the energy leads to unnecessary loss of high frequencies. Second, the source function leads to different energy levels at different frequencies and this calls for different ε_i for each frequency. Finding a single trade-off parameter ε is already challenging, finding one for each frequency requires knowledge on the noise level for each frequency. Third, minimizing energy does not exploit multidimensional structure exhibited by seismic wavefields.

Each of these issues make the inversion of discrete wavefields challenging, especially for noisy real data. To address this challenge, we cast Equation 1 into a form that allows us to solve the unknown wavefield with curvelet-domain sparsity promotion. For this purpose, we use the matrix identity

$$\mathrm{vec}\left(\mathbf{AXB}\right) = \left(\mathbf{B}^T \otimes \mathbf{A}\right) \mathrm{vec}\left(\mathbf{X}\right), \tag{3}$$

which holds for arbitrary matrices $\mathbf{A}, \mathbf{X}$, and $\mathbf{B}$. In this expression $\otimes$ refers to the Kronecker product and vec stands for the linear operation that stacks the columns of a matrix into a long concatenated vector (matlab's colon operator). Equation 1 now becomes

$$\left(\widehat{\mathbf{U}}_i^T \otimes \mathbf{I}\right) \mathrm{vec}\left(\widehat{\mathbf{G}}_i\right) \approx \mathrm{vec}\left(\widehat{\mathbf{V}}_i\right), i = 1 \cdots n_f, \tag{4}$$

with $\mathbf{I}$ the identity matrix. After inclusion of the curvelet synthesis and temporal Fourier transforms ($\mathbf{F}_t = (\mathbf{I} \otimes \mathbf{I} \otimes \mathscr{F}_t)$ with $\mathscr{F}_t$ the temporal Fourier transform), we finally arrive at the following block-diagonal system

$$\underbrace{\mathbf{F}_t^* \begin{bmatrix} \left(\widehat{\mathbf{U}}_1^T \otimes \mathbf{I}\right) & & \\ & \ddots & \\ & & \left(\widehat{\mathbf{U}}_{n_f}^T \otimes \mathbf{I}\right) \end{bmatrix} \mathbf{F}_t}_{\mathbf{U}} \begin{bmatrix} \mathrm{vec}\left(\mathbf{G}_1\right) \\ \vdots \\ \mathrm{vec}\left(\mathbf{G}_{n_t}\right) \end{bmatrix} \approx \begin{bmatrix} \mathrm{vec}\left(\mathbf{V}_1\right) \\ \vdots \\ \mathrm{vec}\left(\mathbf{V}_{n_t}\right) \end{bmatrix}. \tag{5}$$

This equation is amenable to transform-domain sparsity promotion —i.e., Equation 5 can be written as

$$\mathbf{A}\mathbf{x} \approx \mathbf{b} \tag{6}$$

with $\mathbf{A} := \mathbf{U}\mathbf{C}^*$, where $\mathbf{x}$ is the discrete curvelet representation of $g(t,x_s,x_r)$, $\mathbf{C}$ the curvelet transform, and $\widehat{\mathbf{v}}$ the discrete representation of $v(t,x_s,x_r)$.

Solving for the transform-domain representation of the wavefield $g(t,x_s,x_r)$ with t time, x_s the source and x_r the receiver coordinates, now corresponds to inverting a linear system of equations where the monochromatic wavefield $\{\widehat{\mathbf{U}}_i\}_{i=1\cdots n_f}$ and temporal wavefield $\{\mathbf{V}_i\}_{i=1\cdots n_t}$ are related —through the temporal Fourier transform—to the curvelet representation of the discretized wavefield vector $\mathbf{g}$ in the physical domain.

The linear operator $\mathbf{A}$ first applies the inverse curvelet-transform to the vector $\mathbf{x}$, yielding the discrete approximation of the wavefield $g(t,x_s,x_r)$, followed by the application of the temporal Fourier transform operator $\mathbf{F}_t$ and the "vectorized" matrix-matrix multiplication by $\left(\widehat{\mathbf{U}}_i^T \otimes \mathbf{I}\right)$ for each frequency, i.e., for $i = 1\cdots n_f$. In the time domain, these operations correspond to a multidimensional convolution between discrete approximations of the wavefields $g(t,x_s,x_r)$ and $u(t,x_s,x_r)$. Application of $\mathbf{A}^*$ is similar, except that it applies $\left(\widehat{\bar{\mathbf{U}}}_i \otimes \mathbf{I}\right)$ for each frequency (with the bar denoting complex conjugate), followed by the forward curvelet transform. This forward transform is applied to a vector that contains the result of a multidimensional correlation of two discrete wavefields.

EPSI

As shown by Lin et al. (2010); Lin and Herrmann (2011), the above formulation specializes in the case of estimation of primaries into

$$\mathbf{U} := \mathbf{F}_t^* \,\text{blockdiag}\left[\widehat{\mathbf{Q}} - \widehat{\mathbf{P}}\right]_{1\cdots n_f} \mathbf{F}_t, \tag{7}$$

with $\widehat{\mathbf{Q}} = \mathbf{I}\widehat{\mathbf{q}}(\omega)$ the temporal Fourier transform of the source function and $\widehat{\mathbf{P}}$ the Fourier representation of the upgoing wavefield. Throughout this paper, we will assume this source function to be known. In this expression, matrices representing the source functions are full or close to full rank (e.g., $\widehat{\mathbf{Q}} = \mathbf{I}\widehat{\mathbf{q}}(\omega)$ is full rank) and the data is typically rank deficient.

To overcome this rank deficiency, we regularize the inversion by exploiting sparsity by solving

$$\begin{cases} \widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \le \sigma \\ \widetilde{\mathbf{g}} = \mathbf{S}^*\widetilde{\mathbf{x}} \end{cases} \tag{8}$$

in which σ is a noise-dependent tolerance level.We obtain the final solution by applying the synthesis matrix $\mathbf{S}^*$ to the vector $\widetilde{\mathbf{x}}$ (the symbol $\widetilde{}$ denotes a vector obtained by nonlinear optimization) that solves equation 8. To keep things simple for now, we set $\mathbf{S} = \mathbf{I}$ to the identity basis.

DIMENSIONALITY REDUCTION VIA SVD'S

Solving optimization problems (cf. Equation 8) requires multiple iterations involving the application of $\mathbf{A}$, $\mathbf{A}^*$, and possibly $\mathbf{A}^*\mathbf{A}$. In real applications, application of these matrix-vector multiplies is challenging because (i) the matrices are full and extremely large, e.g. for each frequency the data matrix becomes easily $10^6 \times 10^6$ for $n_s = n_r = 1000$ (with n_s the number of sources and n_r the number of receivers), so they require lots of storage and computational resources; (ii) data is incomplete, which requires 'on-the-fly' interpolations that are costly but have the advantage that the data matrix does not need to be formed; and (iii) the solvers require multiple evaluations of $\mathbf{A}$, $\mathbf{A}^*$, and possibly $\mathbf{A}^*\mathbf{A}$. To reduce the storage and multiplication costs of these operations, we replace the data matrix $\widehat{\mathbf{P}}$ by a low-rank approximation.

Because data matrices are large and expensive to form, we adapt a two-stage randomized SVD proposed by Halko et al. (2011). This technique only requires the action of the data matrix on a small (depending on the numerical rank of the data matrix) collection of random vectors, followed by QR factorization. As in Halko et al. (2011), we write the first stage as

$$\widehat{\mathbf{Y}} = \widehat{\mathbf{P}}\widehat{\mathbf{W}}, \tag{9}$$

with $\widehat{\mathbf{W}} \in \mathbb{C}^{n_s \times (k+p)}$ a complex-valued Gaussian random matrix with $k+p$ columns and p a small oversampling factor (typically order $5-10$). So, as long as we can apply the data matrix $k+p$ times to random simultaneous 'shots', we can approximate the action of the full data matrix with a controllable error. For a sufficiently large $k+p$, this error is attained with overwhelming probability (Halko et al., 2011). This means that the matrix $\widehat{\mathbf{Y}}$ contains all information on the range of $\widehat{\mathbf{P}}$ for $k \approx \text{rank}(\widehat{\mathbf{P}})$. As shown by Halko et al. (2011), this allows us to approximate the SVD of $\widehat{\mathbf{P}}$ in three steps:

1. Form a low-rank factorization $\widehat{\mathbf{P}} \approx \mathbf{Q}\mathbf{B}$ with $\mathbf{B} = \mathbf{Q}^*\widehat{\mathbf{P}}$ obtained by a QR-factorization of $\widehat{\mathbf{Y}}$.
2. Compute the SVD of the small matrix $\mathbf{B} = \widetilde{\mathbf{U}}\mathbf{S}\mathbf{V}^*$.
3. Compute $\mathbf{U} = \mathbf{Q}\widetilde{\mathbf{U}}$.

With this we obtain a low k-rank (with $k \ll \min(n_s,n_r)$) approximation of the action of the data matrix via

$$\widehat{\mathbf{P}} \approx \mathbf{U}\mathbf{S}\mathbf{V}^* \tag{10}$$

with $\mathbf{U} \in \mathbb{C}^{n_s \times k}$, $\mathbf{S} \in \mathbb{C}^{k \times k}$, and $\mathbf{V} \in \mathbb{C}^{n_s \times k}$, the dimensionality reduced system. Application of this low-rank approximation leads to faster multiplication by $\widehat{\mathbf{P}}$, $\widehat{\mathbf{P}}^*$, and $\widehat{\mathbf{P}}^*\widehat{\mathbf{P}}$. Using the orthogonality of the left singular vectors, the latter operation can be evaluated efficiently via $\widehat{\mathbf{P}}^*\widehat{\mathbf{P}} \approx \mathbf{V}\mathbf{S}^2\mathbf{V}^*$. Aside from faster matrix multiplications, the above matrix factorization by the SVD also has the advantage that it leads for small k to a significantly reduced memory imprint.

PUTTING EVERYTHING TOGETHER

With EPSI and the low-rank approximation in place, we are now in the position to detail our algorithm to estimate primaries by inverting the low-rank approximation of the data matrix with sparsity promotion. To be successful, we need to strike a balance between memory-reduction and matrix-multiply speedup and the accuracy of the low-rank approximation in the context of sparsity-promoting inversion.

To provide insight in this trade off between accuracy and computational performance, there are three important issues to

consider. First, we need to introduce a norm to measure the accuracy of the low-rank approximation. For this purpose, we introduce aside from the familar Frobenius norm, calculating the two norm of the matrix entries, the spectral/operator norm. The former equals the two-norm of the singular values while the latter is defined by the maximum eigenvalue, i.e. the infinity norm of the eigenvalues. The spectral norm is appropriate because it relates the energy of a vector that is the result matrix-vector product to the energy of the input vector. We denote these two norms by $\|\cdot\|_F$ and $\|\cdot\|_S$. Second, as recently reported by Minato et al. (2011), the numerical rank of the data matrix depends on the frequency as can be observed from Figure 1. We need to incorporate this somehow in our low-rank approximations. Third, the amplitude spectrum of seismic wavelets $|\widehat{\mathbf{q}}(\omega)|$ vary with frequency. Consequently, the energy of the data matrix varies with frequency and this calls for an adaptive scheme to select the rank appropriately.

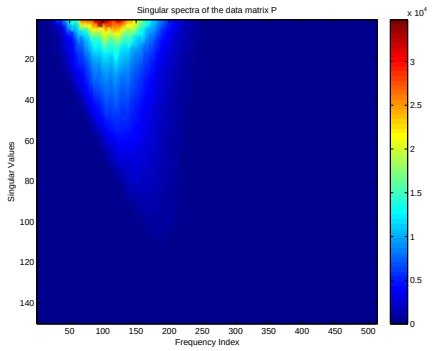

Figure 1: Spectrum of the singular values of the data matrix as a function of frequency. Notice the concentration of energy in the seismic band.

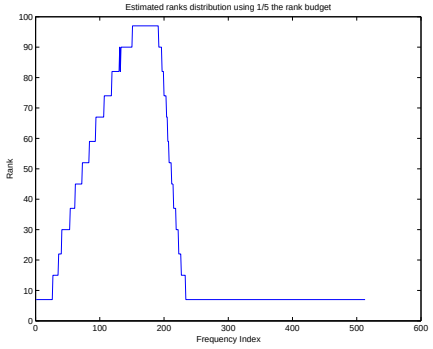

Figure 2: Selected ranks for the low-rank approximation based on a subsampling factor of $\delta = 1/5$. As expected, the a larger rank is selected for the frequencies that are in the seismic band.

To accomplish this 'rank' adaptation, we compare—given a certain total rank budget K—the following two scenarios:

1. assign each frequency with the same rank budget K/n_f and approximate each data matrix with this rank.
2. find for each frequency a rank k such that $\|\widehat{\mathbf{P}} - \mathbf{USV}^*\|_S \leq \varepsilon$ with the accuracy ε fixed and chosen such that the sum over all ranks is smaller than the total rank budget $\sum_\omega k \leq K$. In Figure 2, we include ranks estimates with this procedure. This results clearly shows imprints of the source function and of the fact that the rank of the data matrix increases with frequency because wavefields become more complex for increasing frequency.

PERFORMANCE

To estimate the performance of the proposed algorithm, we compare the output of our robust formulation of EPSI (Lin and Herrmann, 2011) with the output of the second scenario for varying subsampling ratios. For this purpose, we replace the data matrix by its low-rank approximation and use this approximation in EPSI. As expected, the results for the non-adaptive scenario are inferior (13 dB versus 20 dB for $\delta = 1/5$). (Unfortunately, we can not show these results because of space limitations.) The output of EPSI for the adaptive low-rank approximation is included in Figure 3 for a 2D seismic line with $n_s = n_r = 128$ and $n_f = 512$ and $K' = \delta \times \min(n_s, n_r) \times n_f \ll K$. Juxtaposing Figures 3(a) and 3(b) shows relatively little difference between the full data matrix and its low-rank approximation, which can be explained by our adaptation of the rank to the energy (i.e., the seismic wavelet) in the data matrix for each frequency. Similarly, the corresponding estimate for the surface-free Green's function (Figure 3(d)) is also close to the estimate obtained by inverting the surface multiple generator $[\widehat{\mathbf{Q}} - \widehat{\mathbf{P}}]$ with the full data matrix (cf. Figure 3(c)).

To get a better insight in the performance of the algorithm as a function of the subsampling ratio δ, we conducted a series of experiments where we vary the subsampling ratio δ. The results are summarized in table 1 and clearly show that accurate recovery is possible from small subsampling ratios. Ignoring possible slower convergence of the ℓ_1 solver, these subsamplings can lead to significant speedups and reduction of memory imprints.

Of course, there is an up-front cost associated with computing the SVD's. For low-rank matrices, these costs with classical methods are $\mathcal{O}(n_r \times n_s \times K)$ for all data matrices. Using the randomize approach of Halko et al. (2011), these cost are reduced to $\mathcal{O}(n_r \times n_s \times \log K)$, which makes this method potentially computationally feasible for large 3-D problems.

Subsample ratio δ	1/2	1/5	1/8	1/12
	recovery error (dB) / spectral norms ($\times 10^3$)			
	88 (44)	**20** (121)	**16** (144)	**13** (152)
Speed up ($\times$)	2	5	8	12

Table 1: Signal-to-noise ratios, SNR $= 20\log_{10}(\frac{\|\mathbf{g}-\tilde{\mathbf{g}}\|_2}{\|\mathbf{g}\|_2})$ for the estimates of the surface-free Green's function (in bold), and spectral norms (within parenthesis) for different subsample ratios. Notice that the spectral error expressing the difference between the full data matrix and its low-rank approximation increases quasi linearly with the subsampling ratio.

DISCUSSION AND OUTLOOK

The presented method is exciting for the following reasons. First, our approach reduces the storage and multiplication costs by a factor of δ at a small up-front cost of $\mathcal{O}(n_r \times n_s \times \log K)$. This improvement puts us in a good position to scale EPSI to 3-D. Second, the first stage of computing $\widehat{\mathbf{Y}}$ (cf. Equation 9) is equivalent to simultaneous sourcing with random sources (Herrmann et al., 2009). Because this type of sourcing is an instance of compressive sensing, we can use this to further speed

up our algorithm using approaches reported by our group in the literature (see e.g. Herrmann, 2010; Tristan van Leeuwen and Herrmann, 2011). This identification also opens the possibility to work directly with simultaneously acquired Marine data (see e.g. Herrmann, 2010, and another contribution by the authors to these proceedings). Third, the first stage only requires 'black-box' access to the application of the data matrix (read application of SRME (Verschuur et al., 1992; Berkhout and Verschuur, 1997; Weglein et al., 1997)) on a limited number of random simultaneous shots (read random noise). This can be done in parallel and can leverage implementations of the data-matrix multiply that fill in missing data on the fly, which is common practice in 3-D SRME. Again, the burn in costs only require $K+p$ passes through the data, i.e., $K+p$ applications of the SRME operator. Fourth, our low-rank approximations of the data matrix allow us to leverage recent extensions of compressive sensing to matrix completion (Candes and Recht, 2009; Gandy et al., 2011) from incomplete data (read missing traces). In these formulations, data is regularized solving the following optimization problem

$$\widetilde{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{X}\|_* \quad \text{subject to} \quad \|\mathscr{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \sigma, \qquad (11)$$

with $\|\cdot\|_* = \sum |\lambda_i|$ the nuclear norm summing the magnitudes of the singular values (λ) of the matrix $\mathbf{X}$. Here, $\mathscr{A}(\cdot)$ a linear operator that samples the data matrix. It is shown that this program is a convex relaxation of finding the matrix $\mathbf{X}$ with the smallest rank given incomplete data. Low-rank approximations for tensors were recently proposed by Oropeza and Sacchi (2010) for seismic denoising. Fifth, the singular vectors of our low-rank approximation can be used when EPSI is combined with imaging or full-waveform inversion, an approach similar to recent work by Habashy et al. (2010).

CONCLUSIONS

Data-driven methods—such as the estimation of primaries by sparse inversion—suffer from the 'curse of dimensionality' because these methods require repeated applications of the data matrix, whose size grows exponentially with the dimension. In this paper, we leverage recent insights from random matrix theory that allow us to approximate the action of the data matrix via randomized SVDs. The resulting low-rank formulation leads to significant reductions in storage and matrix multiplication costs. The burn-in costs for the low-rank approximations themselves are, by virtue of the randomization, cheap and only require a limited number of applications of the full data matrix to random vectors. This operation can easily be carried out in parallel using existing code bases for surface-related multiple prediction and can lead to significant speedups and reductions in memory use.

ACKNOWLEDGMENTS

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, BP, Chevron, ConocoPhillips, Petrobras, Total SA, and WesternGeco. The first author would like to thank Saudi Aramco for their support.

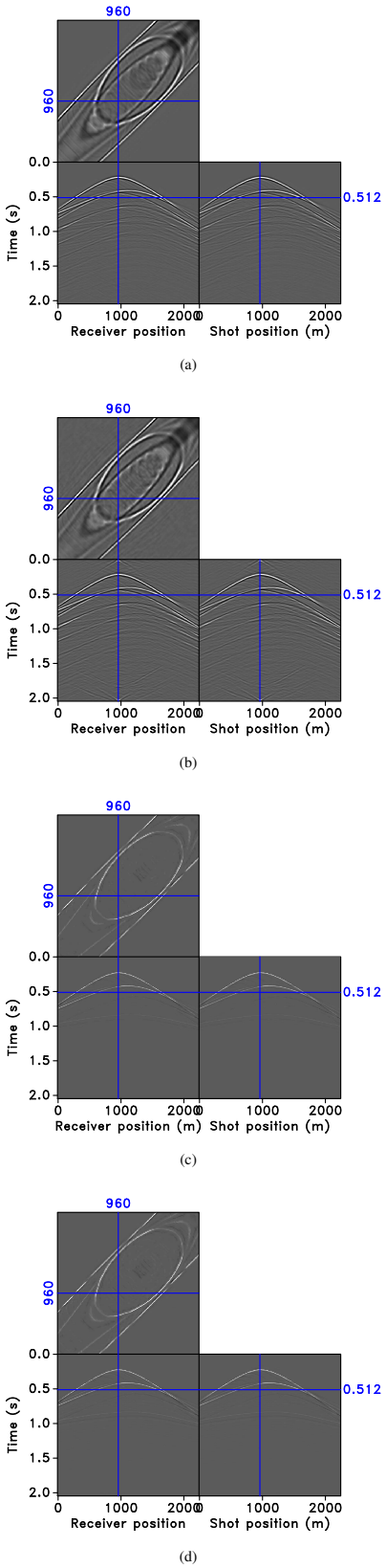

Figure 3: Results of dimensionality-reduced EPSI. **(a)** Input data. **(b)** Low-rank approximation of the data matrix (data in **(a)** transformed into the Fourier domain) with $\delta = 1/5$. **(c)** EPSI result obtained for the full data matrix. **(d)** The same but now for the low-rank approximation. Notice that the results are very close.

REFERENCES

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

Berkhout, A. J., and D. J. Verschuur, 1997, Estimation of multiple scattering by iterative inversion, part I: theoretical considerations: Geophysics, **62**, 1586–1595.

Candes, E., and B. Recht, 2009, Exact matrix completion via convex optimization: Foundations of Computational Mathematics, **9**, 717–772.

Gandy, S., B. Recht, and I. Yamada, 2011, Tensor completion and low-n-rank tensor recovery via convex optimization: Inverse Problems, **27**, 025010.

Habashy, T. M., A. Abubakar, G. Pan, and A. Belani, 2010, Full-waveform seismic inversion using the source-receiver compression approach: , SEG, 1023–1028.

Halko, N., P. G. Martinsson, and J. A. Tropp, 2011, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions: SIAM Review. (to appear June 2011).

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

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

Herrmann, F. J., and D. Wang, 2008, Seismic wavefield inversion with curvelet-domain sparsity promotion: SEG Technical Program Expanded Abstracts, SEG, 2497–2501.

Lin, T., and F. J. Herrmann, 2009, Unified compressive sensing framework for simultaneous acquisition with primary estimation: SEG Technical Program Expanded Abstracts, SEG, 3113–3117.

Lin, T. T., and F. J. Herrmann, 2011, Estimating primaries by sparse inversion in a curvelet-like representation domain: Presented at the , EAGE, EAGE Technical Program Expanded Abstracts.

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

Minato, S., T. Matsuoka, T. Tsuji, D. Draganov, J. Hunziker, and K. Wapenaar, 2011, Seismic interferometry using multidimensional deconvolution and crosscorrelation for crosswell seismic reflection data without borehole sources: Geophysics, **76**, SA19–SA34.

Oropeza, V. E., and M. D. Sacchi, 2010, A randomized svd for multichannel singular spectrum analysis (mssa) noise attenuation: SEG, Expanded Abstracts, **29**, 3539–3544.

Tristan van Leeuwen, A. A., and F. J. Herrmann, 2011, Seismic waveform inversion by stochastic optimization: Technical Report TR-2010-5, UBC-Earth and Ocean Sciences Department. (to appear).

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

Weglein, A. B., F. A. Carvalho, and P. M. Stolt, 1997, An iverse scattering series method for attenuating multiples in seismic reflection data: Geophysics, **62**, 1975–1989.