NUCLEAR NORM MINIMIZATION AND TENSOR COMPLETION IN EXPLORATION SEISMOLOGY

Nadia Kreimer and Mauricio D. Sacchi

Department of Physics, University of Alberta
Edmonton, AB, T6G 2E1, Canada
kreimer@ualberta.ca and msacchi@ualberta.ca

ABSTRACT

We consider the problem of multidimensional seismic data signal recovery and noise attenuation. These data are multidimensional signals that can be described via a low-rank fourth-order tensor in the $frequency-space$ domain. Tensor completion strategies can be used to recover unrecorded observations and to improve the signal-to-noise ratio of seismic data volumes. Tensor completion is posed as an inverse problem and solved via a convex optimization algorithm where a misfit function is minimized in conjunction with the nuclear norm of the tensor. This formulation offers automatic rank determination. We illustrate the performance of the algorithm with a synthetic example and with a real data set obtained by an onshore seismic survey.

Index Terms— seismic data, signal reconstruction, tensor completion, nuclear norm, rank reduction

1. INTRODUCTION

The seismic experiment starts with laying detectors and sources on the surface of a region of interest. Active seismic sources (explosives or vibrators) are utilized to propagate waves in the earth's interior. Arrays of receivers measure wave-fields reflected from subsurface geological boundaries [1, 2]. These data are collected in time series that are usually called seismograms or seismic traces. The acquired data depend on two spatial coordinates for the source and two spatial coordinates for the receivers. These coordinates can be converted to midpoint x, y and offset x, y or offset/azimuth (Fig. 1) coordinates. Modern seismic acquisition systems record large data volumes. These volumes are often irregularly sampled and contain a large number of unrecorded source-receiver positions. Seismic data processing, however, requires fully sampled volumes [3, 4]. The latter is solved via interpolation and reconstruction methods [5]. Reconstruction methods based on signal processing principles can be categorized into: (a) transform-based methods that use the properties of the signal in a suitable domain (e.g. sparsity) [6, 5, 7], (b) Prediction filtering methods that use the predictability of the signal in the *frequency-space* (*F-X*) or *time-space* (*T-X*) domain [8, 9] and (c) rank-reduction methods that utilize the low-rank nature of seismic data [10, 11, 12, 13].

This research has been supported by the sponsors of the Signal Analysis and Imaging Group at the University of Alberta and by a grant to MDS by the Natural Sciences and Engineering Research Council of Canada.

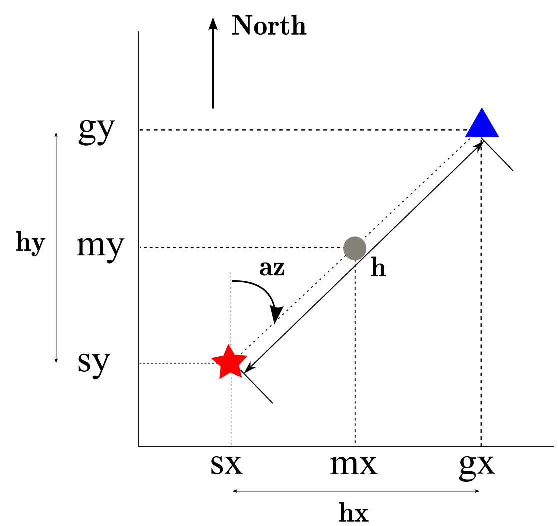

Fig. 1. Definition of coordinates in the seismic experiment. The star is the source, the triangle is the detector (receiver) and the circle is the midpoint location. mx, my denote the midpoint x, y coordinates, hx, hy are the offset x, y coordinates, h is the absolute offset and az is the azimuth angle.

The fully sampled 4D spatial volume has a natural representation via a low-rank tensor structure [13]. Missing traces and noise increase the rank of the tensor. In this context, seismic data reconstruction is equivalent to a tensor completion problem. The completion problem consists of finding a tensor of minimum rank that obeys data constraints. This is an NP-hard problem and a common approach is to replace the rank constraint by the nuclear norm of the tensor [14]. The nuclear norm plays a role similar to the ℓ_1 norm in signal recovery via compressive sensing techniques [15, 16, 14]. Our reconstruction algorithm follows the work of Gandy et al. [17] with a few modifications to make it amenable to reconstructing large volumes of data in the frequency-space domain.

2. THEORY

We will denote tensors with bold calligraphic fonts $\boldsymbol{\mathcal{D}}$, matrices with bold capital fonts $\mathbf{D}$ and scalars with italic letters a. The unfoldings of a fourth-order tensor $\boldsymbol{\mathcal{D}}$ will be written as $\mathbf{D}^{(i)}, i = 1, 2, 3, 4$. Furthermore, the nuclear norm of a matrix $\mathbf{A}$ is $\|\mathbf{A}\|_* = \sum_{i=1}^{n} \sigma_i$, being σ_i the singular values of the matrix. The nuclear norm of a tensor will be defined as the sum of the nuclear norms of its unfoldings. Four unfoldings exist for a fourth-order tensor and the "fold" operation consists of a re-ordering of its elements into a matricial form [13]. The operations of unfolding and folding require careful manipulation of the indices of the tensor and consist of mapping a tensor to a matrix and vice versa.

We assume that a fully sampled seismic volume $\boldsymbol{\mathcal{D}}$ in the *F-X* domain has small n-ranks (rank per unfolding). The cost function for the interpolation and denoising problem is

$$\text{minimize} \quad J = \sum_{i=1}^{4} \|\mathbf{D}^{(i)}\|_* + \frac{\lambda}{2} \|\boldsymbol{\mathcal{T}}\boldsymbol{\mathcal{D}} - \boldsymbol{\mathcal{D}}^{obs}\|_F^2, \quad (1)$$

where $\boldsymbol{\mathcal{D}}^{obs}$ are the observations, $\boldsymbol{\mathcal{T}}$ is the sampling operator with the same size as $\boldsymbol{\mathcal{D}}$ and λ is a trade-off parameter. The minimization of the n-ranks of a tensor is closely related to the truncated Higher-Order singular value decomposition of a tensor [18]. Kreimer and Sacchi [13] proposed to adopt the HOSVD to reconstruct and denoise seismic volumes. In tensor completion via the HOSVD, the user needs to provide the rank of the tensor. An advantage of a nuclear norm formulation, on the other hand, is that the final rank of the tensor is revealed by the optimization algorithm.

The minimization of the objective function in (1) is carried out using the alternating direction method of multipliers (ADMM) [17]. This method solves the problem of minimizing the sum of two convex functions subject to constraints [19]. The minimization of the cost function is carried out one variable at a time, followed by an update of the Lagrange multipliers. In our particular problem (1), we require new variables denoted $\mathbf{Y}_i^{(i)} = \mathbf{D}^{(i)},\ i = 1, 2, 3, 4$ (a tensor $\boldsymbol{\mathcal{Y}}_i$ per unfolding $\mathbf{D}^{(i)}$). We will also have four Lagrange multipliers $\boldsymbol{\mathcal{W}}_i,\ i = 1, 2, 3, 4$.

Using the new split variables $\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{W}}_i$ and identifying the two convex functions to use ADMM [17, 20], the augmented objective function becomes

$$J(\boldsymbol{\mathcal{D}}, \boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{W}}_i) = \frac{\lambda}{2}\|\boldsymbol{\mathcal{T}}\boldsymbol{\mathcal{D}} - \boldsymbol{\mathcal{D}}^{obs}\|_F^2 + \sum_{i=1}^{4} \Big(\|\mathbf{Y}_i^{(i)}\|_* - < \boldsymbol{\mathcal{W}}_i, \boldsymbol{\mathcal{D}} - \boldsymbol{\mathcal{Y}}_i > + \frac{\beta}{2}\|\boldsymbol{\mathcal{D}} - \boldsymbol{\mathcal{Y}}_i\|_F^2 \Big). \quad (2)$$

The ADMM method minimizes J first with respect to $\boldsymbol{\mathcal{D}}$, assuming all other variables constant. Second, it minimizes J with respect to $\boldsymbol{\mathcal{Y}}_i$ for all i while keeping all other variables fixed. The last step is the update of the multipliers $\boldsymbol{\mathcal{W}}_i$.

The minimum of J with respect to $\boldsymbol{\mathcal{Y}}_i$ is the set of the individual minima for each i. Considering the variables $\boldsymbol{\mathcal{D}}, \boldsymbol{\mathcal{W}}_i$ fixed and using the theorem 2.1 from Cai et al. [21], the minimum for $\boldsymbol{\mathcal{Y}}_i$ is

$$\boldsymbol{\mathcal{Y}}_{i\ \min} = \text{fold}\left(\text{shrink}\left(\mathbf{D}^{(i)} - \frac{1}{\beta}\mathbf{W}_i^{(i)}, \frac{1}{\beta}\right)\right), \quad (3)$$

This expression applies for all four variables $\boldsymbol{\mathcal{Y}}_i, i = 1, 2, 3, 4$. The shrinkage operator on a matrix $\mathbf{A}$ is a soft thresholding operator defined as shrink$(\mathbf{A}, \alpha)$ = $\mathbf{U}\tilde{\boldsymbol{\Sigma}}\mathbf{V}^{\mathrm{H}}$, where the singular value decomposition of $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}$ and $\tilde{\boldsymbol{\Sigma}} = \max\{\boldsymbol{\Sigma} - \alpha\mathbf{I}, 0\}$.

To find the minimum with respect to $\boldsymbol{\mathcal{D}}$, we consider the other two variables $\boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{W}}_i$ fixed. The minimum of $\boldsymbol{\mathcal{D}}$ is

$$\boldsymbol{\mathcal{D}}_{\min}\big|_{ijkl} = \quad (4)$$

$$\begin{cases} (\lambda + 4\beta)^{-1} \left[\sum_{i=1}^{4} \boldsymbol{\mathcal{W}}_i + \sum_{i=1}^{4} \beta\boldsymbol{\mathcal{Y}}_i + \lambda\boldsymbol{\mathcal{D}}^{obs} \right] & \text{if } ijkl \in \Omega \\ (4\beta)^{-1} \left[\sum_{i=1}^{4} \boldsymbol{\mathcal{W}}_i + \sum_{i=1}^{4} \beta\boldsymbol{\mathcal{Y}}_i \right] & \text{if } ijkl \notin \Omega \end{cases}$$

where Ω denotes the set of observations. In our algorithm we modified the original method [17] and kept λ, β constant throughout the iterations. These parameters where found using synthetic simulations under different noise and levels of decimation (shown in the following section). Notice that contrary to the implementation in Gandy et al. [17], we apply the ADMM algorithm for each frequency. The convergence proof of this algorithm for constant λ, β can be found in the previously mentioned paper.

Finally, the algorithm reduces to:

repeat until maximum number of iterations k_{max}
    $\boldsymbol{\mathcal{D}}^{k+1}$ as in (4)
for $i = 1, 2, 3, 4$

$$\boldsymbol{\mathcal{Y}}_i^{k+1} = \text{fold}\left(\text{shrink}\left(\mathbf{D}^{(i)\,k+1} - \frac{1}{\beta}\mathbf{W}_i^{(i)\,k}, \frac{1}{\beta}\right)\right)$$

$$\boldsymbol{\mathcal{W}}_i^{k+1} = \boldsymbol{\mathcal{W}}_i^{k} - \beta\left(\boldsymbol{\mathcal{D}}^{k+1} - \boldsymbol{\mathcal{Y}}_i^{k+1}\right)$$

end
    output is $\boldsymbol{\mathcal{D}}^{k_{max}}$

3. SYNTHETIC EXAMPLES

Our synthetic examples are based on a 3D model with two dipping planes. The size of the tensor per frequency is $12 \times 16 \times 12 \times 16$ in the $m_x - m_y - h_x - h_y$ domain. The volume has 512 time samples with a time sampling rate of 2 ms. Therefore, the data has 512 frequencies samples of

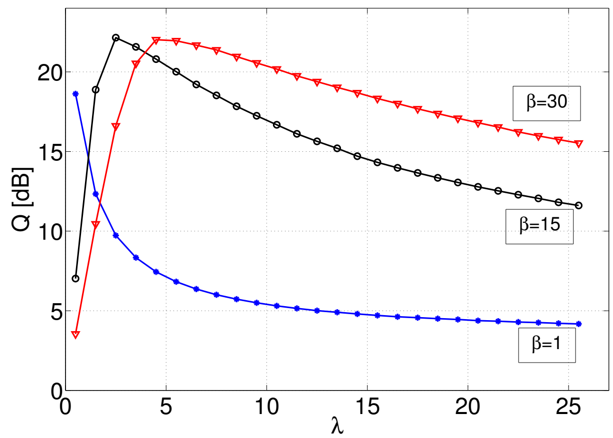

Fig. 2. Quality of the reconstruction Q versus parameter λ for $\beta = 1, 15, 30$.

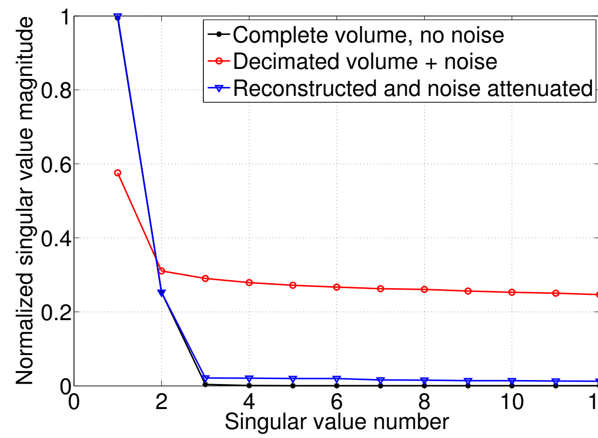

Fig. 3. Normalized singular value distribution for one frequency. Each curve is an average of the singular values for the four unfoldings of the 4D tensor.

which less than half are used for tensor completion. Additionally, we randomly remove 50% of the traces and add random Gaussian noise to produce a volume with SNR $= 1$, where SNR $= \frac{\sigma^2_{\text{data}}}{\sigma^2_{\text{noise}}}$. This synthetic example enables us to select the parameters λ, β that give the optimal results for the presented algorithm and to analyze its behaviour. The quality of the reconstruction in decibel units is $Q = 10 \log \frac{\|\boldsymbol{\mathcal{D}}^{\text{true}}\|^2}{\|\boldsymbol{\mathcal{D}}^{\text{out}} - \boldsymbol{\mathcal{D}}^{\text{true}}\|^2}$, where $\boldsymbol{\mathcal{D}}^{\text{true}}$ is the complete and noise-free volume, and $\boldsymbol{\mathcal{D}}^{\text{out}}$ is the reconstructed and noise-attenuated volume. We use a maximum of 200 iterations for each frequency. We must stress that we are analyzing the global behaviour of the parameters λ, β, .i.e. for all frequencies.

Fig. 2 shows the quality Q versus the trade-off parameter λ for different values of β. We notice that Q deteriorates for the smallest value of β and that a few combinations of λ, β can give similar values of Q. We did not try to use larger values of β because in our tests we observed that increasing this parameter deteriorates the reconstruction. From this particular test, we conclude that larger values of β and smaller values of λ give the best reconstruction. The combination $\lambda = 2.5, \beta = 15$ gives the largest Q and we will use these values throughout the rest of our calculations. An alternative to our choice of β can be found in Boyd et al. [20, p. 20].

The algorithm presented in this paper provides automatic rank determination. In other words, it is not necessary to choose a rank, as in the algorithm proposed by Kreimer and Sacchi [13]. Fig. 3 displays the distribution of singular values for one frequency in the reconstruction, averaged over all unfoldings. We confirm that the reconstructed volume has a similar singular value distribution to the original volume.

With the optimal values of λ, β that were found via the simulations, we reconstruct a 5D synthetic volume of size $512 \times 19 \times 20 \times 19 \times 20$. We remove 40% of the traces randomly and add random Gaussian noise with a SNR $= 1$. Fig. 4 contains a small portion of the volume for two spatial coordinates fixed. The quality of the reconstruction for this case is $Q = 20$ dB. It is important to notice that the waveforms contain significant curvature, which is a problem for other algorithms, such as Fourier based reconstruction methods that use sparsity in the wave-number domain [5, 22]. Evidently, this is not a problem for our algorithm (Fig. 4).

We perform tests to quantify the difference between the HOSVD-based reconstruction algorithm [13] and the presented ADMM-based technique. We apply both methods on the same synthetic example used in Fig. 2 with samples randomly removed from 10% to 90%. We use a maximum of 200 iterations per frequency for the ADMM-based tensor completion and 20 iterations per frequency for the HOSVD-based tensor completion. Fig. 5 contains the results from the simulation. We observe that the ADMM-based method gives better reconstruction results than HOSVD for different levels of decimation at the cost of larger running times. The average running time for ADMM is 45 min for the examples of Fig. 5 on a single processor Intel Xeon(R) running at 3.07 Ghz using MATLAB, while HOSVD's average running time is 2 min.

4. FIELD DATA EXAMPLE

Our real data example is from a land data survey from Alberta, Canada. The grid size utilized is $16 \times 16 \times 8 \times 36$ in the midpoint x, y-offset-azimuth domain. The size of the cells are 10 m$\times$10 m for midpoint x, y directions, 100 m for offset and $10°$ for azimuth. The offset range is $50 - 750$ m. The total number of grid points are $16 \times 16 \times 8 \times 36 = 73728$ whereas the total number of traces are 16306. Therefore, roughly 22% of the grid is populated and our aim is to recover the remaining 78% of the traces. The frequencies considered in the algorithm range from $0.1 - 100$ Hz, with 1 ms sampling rate. As mentioned in the previous section, $\lambda = 2.5$, $\beta = 15$ and

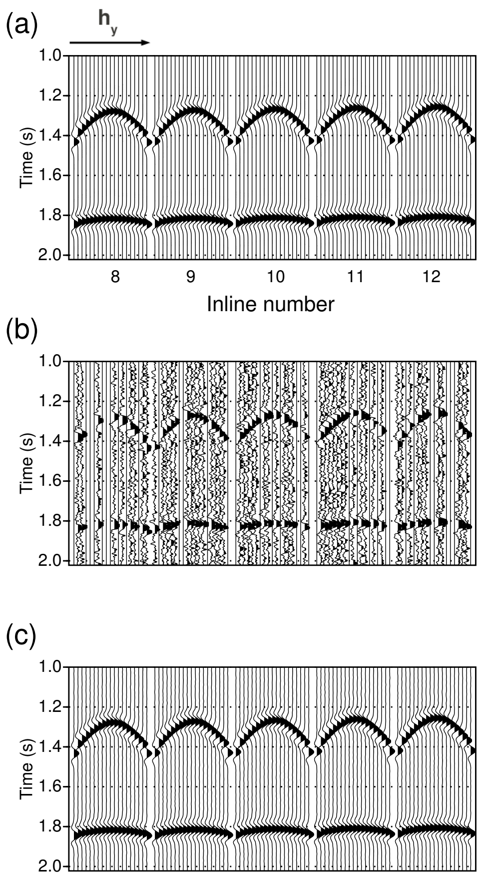

Fig. 4. Small portion of the data with 40% randomly decimated traces in the whole 5D volume and random Gaussian noise with a SNR $=$ 1. (a) Portion of the 5D desired volume. (b) Decimated and noisy data. This is the input to the algorithm. (c) Reconstructed data.

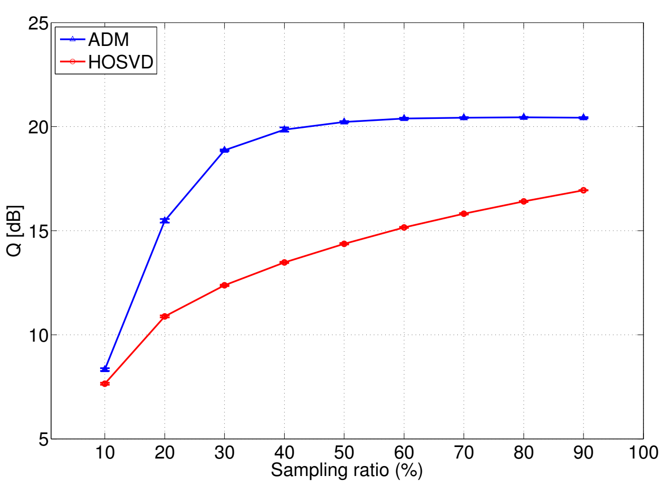

Fig. 5. Quality of the reconstruction versus sampling ratio for a synthetic volume with two different methods. As reference, the largest error bar is equal to 0.1.

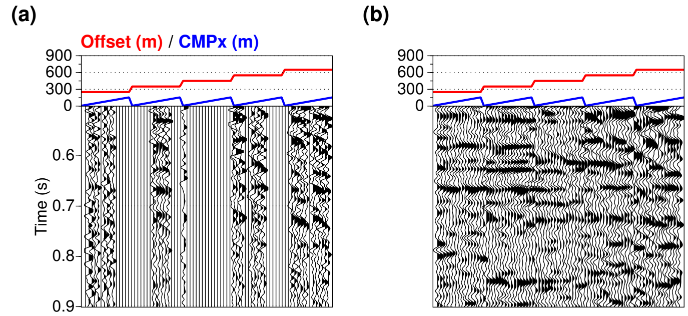

Fig. 6. Subset of the whole volume is displayed in the figure, for two fixed spatial coordinates. a) Input to the algorithm. b) Reconstructed data.

the maximum number of iterations is 200. The running time is 1 h 25 min. Fig. 6 presents a portion of the data for two fixed coordinates and exhibits a satisfactory performance in the reconstruction and noise attenuation of the traces. Naturally, we cannot use the same measure of quality Q as we did for the synthetic data. Although not shown in this article, we can compare the seismic stacks [1] (sets of average of traces falling on the same midpoint bin) before and after to assess the quality of the reconstruction.

5. CONCLUSIONS

This paper focuses on the inverse problem of seismic data reconstruction and denoising. We used the alternating direction method of multipliers one frequency at a time to minimize the cost function of the problem. Unlike other multidimensional reconstruction techniques, the proposed method is tensor completion based and uses nuclear norm minimization of the 4D tensor in the F-X domain. This formulation offers an automatic rank determination of the reconstructed data, bypassing the selection of the rank by the user. Simple simulations allowed us to tune the trade-off parameters. This leads to a high-quality reconstruction even in the presence of seismic events with strong curvature and low SNR. While the assumptions are similar to those of HOSVD-based reconstruction, we observed that this method leads to slightly better results. A real case scenario with a land data example further demonstrates the performance of the algorithm. Although the running times for the proposed algorithm are longer than for HOSVD, ADMM-based reconstruction presents a formal formulation to the rank-reduction based interpolation problem. This formulation contains similarities to current research being done in compressive sensing.

6. REFERENCES

[1] O. Yilmaz, *Seismic Data Processing*, Investigations in Geophysics, Vol 2. Society of Exploration Geophysicists, 1987.

[2] B. L. Biondi, *3D Seismic Imaging*, Society of Exploration Geophysicists, Tulsa, OK, 2006.

[3] M.D. Sacchi and B. Liu, "Minimum weighted norm wavefield reconstruction for AVA imaging," *Geophysical Prospecting*, vol. 53, no. 6, pp. 787–801, 2005.

[4] L. Hunt, J. Downton, S. Reynolds, S. Hadley, D. Trad, and M. Hadleyi, "The effect of interpolation on imaging and AVO: A Viking case study," *Geophysics*, vol. 75, no. 6, pp. WB265–WB274, 2010.

[5] B. Liu and M. D. Sacchi, "Minimum weighted norm interpolation of seismic records," *Geophysics*, vol. 69, no. 6, pp. 1560–1568, 2004.

[6] M. D. Sacchi, T. J. Ulrych, and C. J. Walker, "Interpolation and Extrapolation Using a High-Resolution Discrete Fourier Transform," *IEEE Transaction on Signal Processing*, vol. 46, no. 1, pp. 31–38, 1998.

[7] G. Hennenfent, L. Fenelon, and F. J. Herrmann, "Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach," *Geophysics*, vol. 75, no. 6, pp. WB203–WB210, 2010.

[8] S. Spitz, "Seismic trace interpolation in the *F-X* domain," *Geophysics*, vol. 56, no. 6, pp. 785–794, 1991.

[9] M. Naghizadeh and M. D. Sacchi, "Multistep autoregressive reconstruction of seismic records," *Geophysics*, vol. 72, no. 6, pp. V111–V118, 2007.

[10] S. Trickett, L. Burroughs, A. Milton, L. Walton, and R. Dack, "Rank-reduction-based trace interpolation," *SEG, Expanded Abstracts*, vol. 29, no. 1, pp. 3829–3833, 2010.

[11] V. Oropeza and M.D. Sacchi, "Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis," *Geophysics*, vol. 76, no. 3, pp. V25–V32, 2011.

[12] J. Gao, M.D. Sacchi, and X. Chen, "A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions," *Geophysics*, vol. 78, no. 1, pp. V21–V30, 2013.

[13] N. Kreimer and M.D. Sacchi, "A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation," *Geophysics*, vol. 77, no. 3, pp. V113–V122, 2012.

[14] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proceedings of the American Control Conference*, 2001, pp. 4734–4739.

[15] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[16] E.J. Candès and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925 –936, june 2010.

[17] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Problems*, vol. 27, no. 2, pp. 025010, 2011.

[18] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[19] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice Hall, 1989.

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[21] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[22] D. Trad, "Five-dimensional interpolation: Recovering from acquisition constraints," *Geophysics*, vol. 74, no. 6, pp. V123–V132, 2009.