# 5D and 4D Pre-stack seismic data completion using tensor nuclear norm (TNN)

*Gregory Ely, Shuchin Aeron, Ning Hao, & Misha E. Kilmer*

## SUMMARY

In this paper we present novel strategies for completion of 5D pre-stack seismic data, viewed as a 5D tensor or as a set of 4D tensors across temporal frequencies. In contrast to existing complexity penalized algorithms for seismic data completion, which employ matrix analogues of tensor decompositions such as HOSVD or use overlapped Schatten norms from different unfoldings (or matricization) of the tensors, our approach uses a recently proposed decomposition called tensor SVD or tSVD for short, proposed in [Kilmer and Martin (2011)]. We show that seismic data exhibits low complexity under tSVD, i.e. is compressible under tSVD representation, and we subsequently propose a new complexity penalized algorithm for pre-stack seismic data completion under missing traces. This complexity measure which we call the Tensor Nuclear Norm (TNN) is motivated by algebraic properties of the tSVD. We test the performance of the proposed algorithms on synthetic and real data and show that missing data can be reliably recovered under heavy down-sampling.

## INTRODUCTION

This paper will mainly discuss recovery of pre-stack seismic records from incomplete spatial measurements. Ideally, a seismic survey would have receiver and source geometries at all of the possible spatial indices, however this rarely occurs due to financial and physical constraints. Instead, only a fraction of the possible shot and receiver locations are recorded and the data is spatially under-sampled, confounding interpretation of the seismic data, see [Trad (2009)] and references therein.

The pre-stack seismic data can be viewed as a 5D data or a fifth order *tensor* consisting of one time or frequency dimension and four spatial dimensions describing the location of the detector and the receiver in a two dimensional plane. This data can then be described in terms of the original $(r_x, r_y, s_x, s_y)$ coordinate frames or in terms of midpoint receivers and offsets $(x, y, h_x, h_y)$. Alternatively, one can also view this data as a collection of 4D tensors across the temporal frequencies as considered in [Trad (2009)] and [Kreimer and Sacchi (2011)].

In this paper we present a novel method for seismic data completion from the limited survey information using complexity penalized recovery algorithms which measure complexity under the appropriate tensor decompositions. We assume that the seismic data has low information complexity in the sense that there is high redundancy or correlation in the traces from one grid point to the next. These tensor complexity measures capture this redundancy across the traces. From the perspective of optimal sampling and recovery, low complexity data can be reliably recovered from a measurement rate in proportion to the information rate of the data, which is a generic theme in compressive sensing theory and methods [Donoho (2006); Candes and Tao (2006); Aeron et al. (2010)]. In the context of data viewed as a matrix (or a $2^{nd}$ order tensor) one complexity measure that has shown promise in matrix completion problems is matrix rank [Candès and Recht (2012)] (computed via its Singular Value Decomposition (SVD)). The notion of rank extends to higher order tensors to capture the redundancy of the data but unlike the 2D case, for higher order tensors rank depends highly on the type of decomposition or factorization. Many methods exist for tensor decompositions together with their associated notions of rank, primary among them being the Higher Order SVD (HOSVD) and Canonical Decomposition (CANDECOMP), see [Kolda and Bader (2009)] for a survey of these decompositions. Other notions are based on flattening of the tensor into matrices and then computing the ranks of the resulting matrices, e.g. the tensor $p$-rank of order $p$ tensors which is the sum of ranks of the $p$ matrices obtained by flattening the tensor along the $p$ different directions [ Gandy et al. (2011)].

Seismic data completion using an appropriate convex relaxation of these complexity measures have been proposed recently in [Kreimer and Sacchi (2011, 2012b)], [Gao et al. (2011)] and [Kumar et al. (2013); Silva and Herrmann (2013)].In this paper we exploit a novel tensor singular value decomposition (tSVD) proposed in [Kilmer and Martin (2011)] for third order tensors and its extensions to higher order tensors in [Martin et al. (2013)]. With synthetic data we show that the pre-stack seismic data is compressible, i.e. has low informational complexity in the tSVD domain and hence can be reliably recovered under limited sampling using an appropriate complexity penalized algorithm in the tSVD domain. The following section describes mathematical preliminaries to state the problem and background needed for understanding the proposed algorithms.

## METHOD

The true seismic data $\mathcal{M}$ is sparially under-sampled and this operation can be represented by a linear operator $\mathcal{A}$ resulting in the sparse observed data $\mathcal{Y}$ under additive noise.

$$\mathcal{Y} = \mathcal{A}(\mathcal{M}) + \mathcal{N} \tag{1}$$

The problem of seismic data completion becomes to reliably estimate $\hat{\mathcal{M}}$ from $\mathcal{Y}$ under the sampling operator $\mathcal{A}$. However, because the number of observed measurements is significantly less than the number of elements in $\mathcal{M}$, the problem is severely ill-posed and cannot be solved directly without placing some constraints on $\mathcal{M}$. As noted in the introduction, these constraints arise in the form of complexity measures on the underlying true data $\mathcal{M}$ and one can reliably recover the $\mathcal{M}$ using the complexity penalized algorithms of the type,

$$\begin{aligned} &\min\ h(\mathcal{M}) \\ &s.t. ||\mathcal{Y} - \mathcal{A}(\mathcal{M})|| \leq \sigma_n \end{aligned} \tag{2}$$

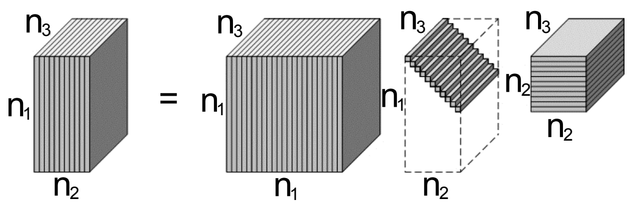


Figure 1: The t-SVD of an $n_1 \times n_2 \times n_3$ tensor. A tensor can be regarded as a matrix of fibers or tubes along the third dimension of a tensor $\mathcal{M}$. Then tSVD is analogous to a matrix SVD if we regard the diagonal tensor $\mathcal{S}$ as consisting of singular "tubes" or "vectors" on the diagonal analogous to singular values on the diagonal in the traditional SVD. For tensors of order $p$ tSVD extends the notion of singular value to higher dimensions, in which each *tube* can be represented as $p-1$ dimensional tensor. For example, a 4D tensor of size $n_1 \times n_2 \times n_3 \times n_4$ has a tSVD decomposition in which each *tubal* singular value is a 3D tensor of size $n_1 \times 1 \times n_3 \times n_4$.

where $h(\mathcal{M}) \in \mathbb{R}^+ \cup \{0\}$ is a non-negative real valued mapping and measures the complexity of the true data $\mathcal{M}$ and $\sigma_n^2$ is the additive noise variance. In the following, we assume that the noise variance is negligible. Since our complexity measure derived from tSVD, in the next section we provide a brief overview of the decomposition.

**Math Background: tSVD**

Herein we provide a brief overview of the tSVD framework. For more details, please refer to [Kilmer and Martin (2011)]. [Kilmer et al. (2013)] and [Martin et al. (2013)]. For the sake of visual and analytical interpretation, we discuss the 3D case here briefly. The 3D tSVD extends in a recursive manner to higher order tSVD, as shown in [Martin et al. (2013)]. Let $\mathcal{M}$ be an order $p$ tensor $\in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_p}$, then $\mathcal{M}$ can be decomposed as the following form,

$$\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}} \tag{3}$$

where $\mathcal{U}$ is an *orthogonal* tensor of size $n_1 \times n_1 \times n_3 \times \ldots \times n_p$, $\mathcal{S}$ is a block-diagonal tensor of size $n_1 \times n_2 \ldots \times n_p$, $\mathcal{V}$ is an *orthogonal* tensor of size $n_2 \times n_2 \times n_3 \times \ldots \times n_p$ and $*$ denotes the t-product. The orthogonality of $\mathcal{U}$ and $\mathcal{V}$ is understood in terms of the t-product and a tensor transpose operation, namely $\mathcal{U} * \mathcal{U}^T = \mathcal{I}$ where $\mathcal{I}$ denotes the identity tensor such that for all tensors of appropriate dimension $\mathcal{M} * \mathcal{I} = \mathcal{M}$, see [Kilmer et al. (2013)] for details. This SVD like decomposition is best understood by looking at Figure 1 where we show the decomposition for 3D case. As explained in [Kilmer and Martin (2011)], due to the nature by which the tensor product $*$ is defined, this tensor factorization can be obtained by **Algorithm 1** using FFT, taken verbatim from [Martin et al. (2013)], where in line with notation and convention there, we have chosen the convention in MATLAB which stores higher order tensors as a stack of matrices with the stack index running from 1 to $\rho = n_3 n_4 \ldots n_5$.

The main advantage of this type of decomposition stems from the optimality property of a "truncated-tSVD" as a best *t-rank* $k$ approximation, where $k \le l = \min\{n_1, n_2\}$, to a tensor in the Frobenius norm where the rank is defined with respect to tSVD

**Algorithm 1** tSVD

**Input:** $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \ldots \times n_p}$
$\rho = n_3 n_4 \ldots n_p$
**for** $i = 3$ to p **do**
  $\mathcal{D} \leftarrow$ fft($\mathcal{M}$, [ ], i);
**end for**
**for** $i = 1$ to $\rho$ **do**
  $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = svd(\mathcal{D}(:,:,i))$
  $\mathcal{U}(:,:,i) = \mathbf{U}$; $\mathcal{S}(:,:,i) = \mathbf{S}$; $\mathcal{V}(:,:,i) = \mathbf{V}$;
**end for**
**for** $i = 3$ to p **do**
  $\mathcal{U} \leftarrow$ ifft($\mathcal{U}$, [ ], i); $\mathcal{S} \leftarrow$ ifft($\mathcal{S}$, [ ], i); $\mathcal{V} \leftarrow$ ifft($\mathcal{V}$, [ ], )i;
**end for**

as the number of non-zero singular "tubes" in SVD, [Kilmer and Martin (2011)]. This optimality results from the natural ordering of the singular tubes by their energy captured by the $\ell_2$ norms of the tubal vectors. This key feature distinguishes tSVD from HOSVD[1] where there is no natural ordering of the entries in the core tensor. Other tensor analysis methods, such as tensor $p$-rank used in [Gandy et al. (2011)] , operate by flattening the tensor and do not respect the *orientation* of the tensor and the structural complexity aspects inherent in the orientation are lost. In contrast, the tSVD based approach preserves this orientation and can be taken along the most natural orientation of the data.

In this context, tensor completion using tSVD exploits the compressibility of natural data in the tSVD domain in the sense that it can be well approximated by a low-rank tSVD, where the notion of the rank under tSVD is presented in [Kilmer et al. (2013)]. There the *multi-rank* of a tensor using tSVD was defined to be a *vector of the ranks* of the frontal slices $\mathcal{D}(:,:,i)$, see **Algorithm 1**. This suggests a complexity penalized algorithm for recovery of $\mathcal{M}$ from linear measurements, that minimizes the sum of the elements of the tubal-rank vector of $\mathcal{M}$.

$$\begin{aligned} \mathrm{CP}: \ \min \sum_{i=1}^{\rho} \mathrm{rank}(\mathcal{D}(:,:,i)) \\ s.t. \mathcal{Y} = \mathcal{A}(\mathcal{M}) \end{aligned} \tag{4}$$

However, as in the case of recovery of matrices under-sampling via linear operators by minimizing rank, the problem CP is NP-hard. We therefore *relax* the complexity measure to a norm which we call the tensor nuclear norm (TNN) [Semerci et al. (2012)],

$$\begin{aligned} \mathrm{OPT\_TNN}: \min \sum_{i=1}^{\rho} ||\mathcal{D}(:,:,i)||_{nuc} \\ s.t. \ \mathcal{Y} = \mathcal{A}(\mathcal{M}) \end{aligned} \tag{5}$$

where $||\cdot||_{nuc}$ denotes the Schatten-1 norm on the singular values of the matrix in the argument [Watson (1992)], also known as the nuclear norm in the literature [Candès and Recht (2012)]. It is easy to see that the proposed optimization problem is a convex optimization problem, see also [Semerci et al.

[1] HOSVD is the only reasonably computable decompositions for higher order tensors. CANDECOMP may or may not be unique generally and when it is unique finding the right decomposition is usually computationally intensive.

(2012)] for details, and therefore can be solved using existing techniques. Before we present the algorithm, in the next section we first show that the pre-stack seismic data is compressible in the tSVD domain.

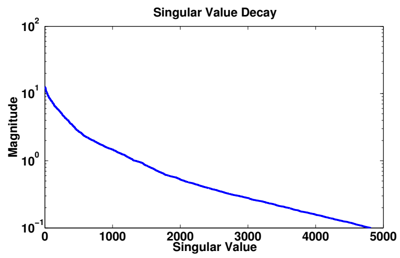


Figure 2: This figure shows the decay of singular values of the synthetic seismic data which empirically obey a power law decay.

**Compressibility of seismic data in the tSVD domain**

In order to demonstrate the compressibility of the seismic data, we generated a synthetic 5D survey in which sources and receivers were placed on a $12 \times 12$ grid spaced 100 meters apart. Three synthetic Born scatters were placed below the surface and traces consisting of $N_t$ samples were generated for all of the possible receivers source geometry to generate a 5D tensor with dimensions $12 \times 12 \times 12 \times 12 \times N_t$. Several $r_y$ directional slices of the original tensor are shown in Figure 4 for different values $r_x$ with source location being held constant at $s_x$ 3 and $s_y$ 7. The tSVD was applied to the synthetic data. Figure 2 shows the decay of the singular values of the matrices $\mathcal{D}(:,:,i)$ as computed using the tSVD and for the synthetic data case they obey a power law decay, which implies that $\sum_{i=1}^{\rho} ||\mathcal{D}(:,:,i)||_{nuc}$ is sufficient measure of complexity of the seismic data in the tSVD domain. As a result, we expect the data to be recoverable from a limited number of measurements by using the algorithm OPT_TNN.

**A first order algorithm for solving OPT_TNN**

Motivated by recently proposed first order iterative shrinkage algorithms for solving nuclear norm minimization problems [Cai et al. (2010)], we employ **Algorithm 2** for solving OPT_TNN. Where $Sh_\varepsilon$ is an element-wise shrinkage function that applies a soft thresholding to $\mathcal{S}$ in the Fourier domain, see Equation 6. $\varepsilon$ determines the step size and $\alpha$ controls the tolerance of the equality constraint. For all of the instances of the algorithm, a value of 1 & 1 was chosen for these two parameters and the algorithm ran for 100 iterations. It is observed that the result is robust to the choice of these parameters although for some extreme values the algorithm performance degrades.

$$Sh_\varepsilon[x] = \begin{cases} x-\varepsilon, \text{ if } x > \varepsilon, \\ x+\varepsilon, \text{ if } x < \varepsilon, \\ 0, \text{otherwise}, \end{cases} \quad (6)$$

## PERFORMANCE EVALUATION: SYNTHETIC DATA

To evaluate our algorithm we used the synthetic dataset described previously and removed 25, 50, 70, 80, 85, and 90 percent of the measured traces, resulting in highly under-sampled data as shown in Figure 4B for 90% of traces removed. **Algorithm 2** was then applied in two ways: (1) Recovering the data by completion of 4D tensors frequency by frequency (2) Recovering the entire 5D seismic volume at once. Since the downsampling operator is chosen at random, we averaged the performance in recovery over 10 randomly selected downsampling operators for each level of under-sampling. The error in recovery is measured via Normalized Root Mean Square Error (NRMSE). For the synthetic data, the performance as function of sampling rate is shown in Figure 3. Although the NRMSE is significant for the highly under-sampled data, the reflections are still highly visible in the reconstruction, Figure 4. Therefore, the NRMSE appears to be a pessimistic norm of the quality of the reconstruction and one should focus on the errors at the time support of the arrivals instead of the entire time window. For highly under-sampled data, the reconstruction using the full 5D data rather than the frequency by frequency data has slightly less NRMSE, as shown in the Figure 3. Unlike the methods considered so far in [Kreimer and Sacchi (2011, 2012b)] which work on 4D data frequency by frequency, we observed a small benefit to processing the data as a whole.

**Algorithm 2** Iterative Shrinkage

$\mathcal{P} = null(\mathcal{A})$ // Projects onto the nullspace of the measurement tensor.
$\mathcal{X} = \mathcal{Z} = \mathcal{B} = 0$ // Initialize variables.
$\rho = n_3 n_4 ... n_p$
**while** Not Converged **do**
  $\mathcal{X} = \mathcal{P} \cdot (\mathcal{Z} - \mathcal{B}) + \mathcal{Y};$
  $\mathcal{Z}_{old} = \mathcal{Z}$
  $\hat{\mathcal{X}} = \alpha \hat{\mathcal{X}} + (1-\alpha)\mathcal{Z}_{old}$
  $\mathcal{M} = \hat{\mathcal{X}} + \mathcal{B}$
  // Calulcate tSVD using Algorithm 1 and apply shrinkage in the tSVD domain.
  **for** $i = 3$ to p **do**
    $\mathcal{D} \leftarrow \text{fft}(\mathcal{M}, [\,], \text{i});$
  **end for**
  **for** $i = 1$ to $\rho$ **do**
    $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = svd(\mathcal{D}(:,:,i));$
    $\mathbf{S} = Sh_\varepsilon[\mathbf{S}];$
    $\mathcal{U}(:,:,i) = \mathbf{U};\ \mathcal{S}(:,:,i) = \mathbf{S};\ \mathcal{V}(:,:,i) = \mathbf{V};$
  **end for**
  **for** $i = 3$ to p **do**
    $\mathcal{U} \leftarrow \text{ifft}(\mathcal{U}, [\,], \text{i});\ \mathcal{S} \leftarrow \text{ifft}(\mathcal{S}, [\,], \text{i});\ \mathcal{V} \leftarrow \text{ifft}(\mathcal{V}, [\,], )\text{i};$
  **end for**
  $\mathcal{Z} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$
  $\mathcal{B} = \mathcal{B} + \hat{\mathcal{X}} - \mathcal{Z}$
**end while**

## PERFORMANCE ON FIELD DATA

In addition to synthetic data, the algorithm was tested on a survey of the Western Canadian Sedimentary Basin. The dataset consists of grid with 29 midpoint gather locations with $12 \times 12$ offset coordinates to generate a data set with spatial dimensions of $29 \times 29 \times 12 \times 12$. The midpoint grid spacing is 26 meters and 52 meters for x and y with an offset spacing from 0 to 1400 meters. Given the geometry roughly 121,000 source

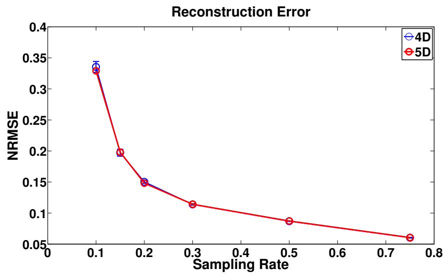


Figure 3: This figure shows the reconstruction error as function of sampling fraction for both the 4D frequency by frequency and full 5D reconstruction. For severely under-sampled data, below 20 percent, the 5D reconstruction provides marginally better results than the 4D reconstruction.

receiver geometries are possible, however only 16060 traces were recorded resulting in sampling rate of approximately 8%. The reconstructed results for the a fixed y-midpoint gather of 11 and y-offset of 6 are shown in figure 5. These traces are the same as in [Kreimer and Sacchi (2012a)] which used HOSVD was used to reconstruct the same dataset.

## CONCLUSIONS & FUTURE WORK

We have presented a novel method for the reconstruction of incomplete seismic through the use of tSVD rank minimization algorithm. The algorithm was tested both in a frequency by frequency manner as well as applied to the full 5D tensor. For highly under-sampled data, completing in the full 5D space results in slightly better reconstructions. Due to the construction of the tSVD and the shrinkage operator in the tSVD domain, both the 5D and 4D have equivalent computational load and therefore the full 5D reconstruction should be used. Note that in this paper we have presented a completion algorithm which is robust in presence of small amounts of noise. In future we will extend this to the noisy case via the following optimization problem,

$$\text{Noisy\_OPT\_TNN}: \min \sum_{i=1}^{\rho} ||\mathcal{D}(:,:,i)||_{nuc} + \gamma ||\mathcal{N}||_2^2$$
$$\text{s.t. } \mathcal{Y} = \mathcal{A}(\mathcal{M}) + \mathcal{N} \tag{7}$$

for some $\gamma > 0$ (regularization parameter) which can also be solved efficiently by first order methods. **Alternatively**, the de-noising after completion can be achieved by thresholding in the discrete curvelet domain, see foe e.g. [Hennenfent and Herrmann (2006)] after completion. This strategy is similar to the thresholded basis pursuit algorithm of [Saligrama and Zhao (2011)] which is also nearly optimal.

## ACKNOWLEDGMENT

We would like to thank CGG, Canada and Prof. Mauricio Sacchi at University of Alberta, Canada for providing us with the field data. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. (DGE- 0806676).

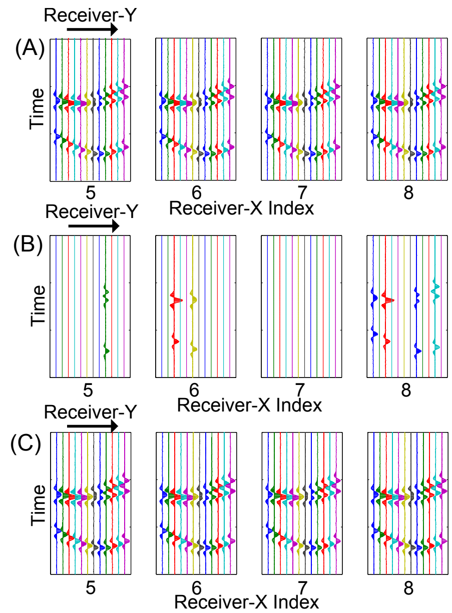


Figure 4: This figure shows the full synthetic data (A) for four different receiver source slices as well as the under-sampled measured data for the case when 90% of the traces were removed (B). In addition, the reconstruction for the 5D (C) slices are shown as well.

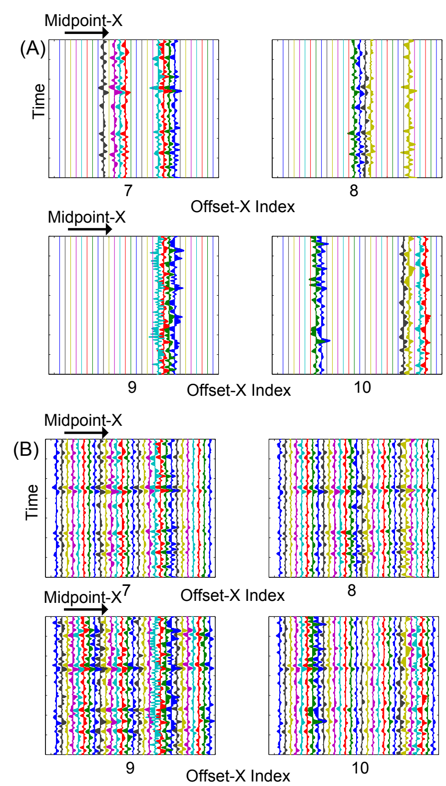


Figure 5: This figure (A) shows the sparsely sampled field data from the Western Canadian Sedimentary Basin and the reconstructed traces (B) using a frequency by frequency procedure.

## REFERENCES


Aeron, S., V. Saligrama, and M. Zhao, 2010, Information theoretic bounds for compressed sensing: IEEE Transactions on Information Theory, **56**, 5111–5130.

Cai, J. F., E. J. Candès, and Z. Shen, 2010, A singular value thresholding algorithm for matrix completion: SIAM Journal on Optimization, **20**, 1956–1982.

Candes, E., and T. Tao, 2006, Near optimal signal recovery from random projections: Universal encoding strategies?: IEEE Transactions on Information Theory, **52**, 5406–5425.

Candès, E. J., and B. Recht, 2012, Exact matrix completion via convex optimization: Commun. ACM, **55**, 111–119.

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

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

Gao, J., M. D. Sacchi, and X. Chen, 2011, A fast rank reduction method for the reconstruction of 5D seismic volumes: SEG Technical Program Expanded Abstracts, 3622–3627.

Hennenfent, G., and F. Herrmann, 2006, Seismic denoising with nonuniformly sampled curvelets: Computing in Science Engineering, **8**, 16–25.

Kilmer, M., K. Braman, N. Hao, and R. Hoover, 2013, Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging: SIAM Journal on Matrix Analysis and Applications, **34**, 148–172.

Kilmer, M. E., and C. D. Martin, 2011, Factorization strategies for third-order tensors: Linear Algebra and Its Applications, **435**, 641–658. (WOS:000291142900013).

Kolda, T., and B. Bader, 2009, Tensor decompositions and applications: SIAM Review, **51**, 455–500.

Kreimer, N., and M. Sacchi, 2012a, Reconstruction of seismic data via tensor completion: Statistical Signal Processing Workshop (SSP), 2012 IEEE, 29–32.

Kreimer, N., and M. D. Sacchi, 2011, A tensor higher order singular value decomposition (hosvd) for pre-stack simultaneous noise reduction and interpolation, *in* SEG Technical Program Expanded Abstracts 2011: SEG, 3069–3074.

———, 2012b, Tensor completion via nuclear norm minimization for 5d seismic data reconstruction, *in* SEG Technical Program Expanded Abstracts: SEG, 1–5.

Kumar, R., A. Y. Aravkin, H. Mansour, B. Recht, and F. J. Herrmann, 2013, Seismic data interpolation and denoising using svd-free low-rank matrix factorization. https://www.slim.eos.ubc.ca/Publications/Public/Conferences/EAGE/2013/kumar2013EAGEsind/kumar2013EAGEsind.pdf.

Martin, C. D., R. Shafer, and B. LaRue, 2013, An order-$p$ tensor factorization with applications in imaging: SIAM Journal on Scientific Computing, **35**, A474–A490.

Saligrama, V., and M. Zhao, 2011, Thresholded basis pursuit: Lp algorithm for order-wise optimal support recovery for sparse and approximately sparse signals from noisy random measurements: IEEE Transactions on Information Theory, **57**, 1567–1586.

Semerci, O., N. Hao, M. Kilmer, and E. Miller, 2012, Tensor-based formulation for spectral computed tomography: Presented at the SIAM Conference on Imaging Science.

Silva, C. D., and F. J. Herrmann, 2013, Hierarchical tucker tensor optimization - applications to 4d seismic data interpolation. https://www.slim.eos.ubc.ca/Publications/Public/Conferences/ EAGE/2013/dasilva2013EAGEhtucktensor/dasilva2013EAGEhtucktensor.pdf.

Trad, D., 2009, Five-dimensional interpolation: Recovering from acquisition constraints: GEOPHYSICS, **74**, V123–V132.

Watson, G., 1992, Characterization of the subdifferential of some matrix norms: Linear Algebra and its Applications, **170**, 33 – 45.