Theoretical Performance Analysis of Tucker Higher Order SVD in Extracting Structure from Multiple Signal-plus-Noise Matrices

Himanshu Nayar
Dept. of EECS,
University of Michigan,
Ann Arbor, Michigan, 48104
email: hnayar@umich.edu

Raj Rao Nadakuditi
Dept. of EECS,
University of Michigan,
Ann Arbor, Michigan, 48104
email: rajnrao@umich.edu

***Abstract*—The Tucker Higher Order SVD is a popular algorithm for uncovering structure from tensor datacubes. This algorithm has been successfully used in many signal processing, machine learning and data mining applications. In this work, we use recent results from random matrix theory to analyze the performance of the HOSVD algorithm. In particular, we focus on the performance HOSVD on 3-D tensors for extraction of structure from signal-plus-noise tensors. We analyze the missing data setting where the entries of the signal-plus-noise tensor are randomly deleted. Our analysis bring into focus a phase transition phenomenon which separates a regime where the HOSVD can accurately estimate the latent signal matrix from a regime where it cannot. The threshold depends on the signal-to-noise ratio and the fraction of missing entries observed in a manner we make explicit. Finally, we illustrate the predicted performance curves using numerical simulations and illustrate implication of our predictions on an widely used facial recognition dataset.**

I. Introduction

The Tucker HOSVD is a popular algorithm for extracting structure from datacubes [1]–[3]. In the signal processing literature, the HOSVD has been shown to successfully tease out weak signals from multiple datasets that are organized as a tensor or a datacube; see, for e.g., [4]–[8]. The authors of these methods have pointed out the need to rigorously quantify the performance of the HOSVD so that it may be compared to the performance of other non-HOSVD based methods for extracting structure from datacubes [9], [10].

In this work, we consider a setting where we are given n, $m_1 \times m_2$ signal-plus-noise matrices $\widetilde{\mathcal{X}}(:,:,1),\ldots,\widetilde{\mathcal{X}}(:,:,n)$. Motivated by problems arising in EEG analysis, image processing and other such data fusion applications [11]–[15], we model these matrices as

$$\widetilde{\mathcal{X}}(:,:,i) \quad = \sum_{j=1}^{k} \mathbf{v}(i,j)A_{(j)} + \mathcal{Z}(:,:,i), \text{ for } i = 1,\ldots,n,$$

and analyze the performance of the HOSVD algorithm via its ability to recover the latent $A_{(j)}$ matrices. The main contribution of this paper is the first mathematically rigorous characterization of the performance of the HOSVD in extracting underlying latent signal matrices. The results build on recent results from random matrix theory [19], [20]. We expect that this work can form the basis for improve the understanding of the fundamental performance limits of the HOSVD in the context of the many signal processing applications listed earlier. Our results bring into sharp focus a phase transition which separates a regime where the HOSVD recovers correlated estimates of the latent signal matrices from a regime where the HOSVD recovers (asymptotically) uncorrelated estimates of the latent signal matrices. We precisely characterize the phase transition boundary and how it is affected in the missing data setting.

The paper is organized as follows. In Section II, we discuss the signal-plus-noise models for the datacube; we summarize the HOSVD algorithm and the quantity of interest in Section III. We present the main theoretical results in Section IV, provide a sketch of the proof in Section V and corroborate the theoretical results with numerical simulations in Section VI. We present some concluding remarks in Section VII.

II. Signal-plus-noise tensor models

We are given a 3-D tensor $\widetilde{\mathcal{X}}$ whose i^{th} layer $\widetilde{\mathcal{X}}(:,:,i)$ is modeled as

$$\widetilde{\mathcal{X}}(:,:,i) \quad = \sum_{j=1}^{k} \underbrace{\mathbf{v}(i,j)A_{(j)}}_{\mathcal{L}(:,:,i)} + \mathcal{Z}(:,:,i), \quad \text{for } i = 1,\ldots,n, \quad (1)$$

where $\mathbf{v}(i,j)$ are the entries of an arbitrarily chosen matrix $\mathbf{v}$ with orthonormal columns, $A(j) \in \mathbb{R}^{m_1 \times m_2} : j \in \{1,\ldots,k\}$ for $j \in \{1 \ldots k\}$ are arbitrary 2-D matrices such that $\langle A_{(i)}, A_{(j)} \rangle = \text{Tr}(A_{(i)}^T A_{(j)}) = \theta_i \theta_j \delta_{i,j}$ and k is fixed with respect to n. The tensor $\mathcal{L}$ should be considered as the *signal* tensor. Each entry of noise-only tensor $\mathcal{Z}$ is assumed to be i.i.d. Gaussian with mean zero and variance $1/n$. Thus, (1), describes $\widetilde{\mathcal{X}}$ as a signal-plus-noise tensor.

We also consider the missing data setting. Here, we are given a 3-D tensor $\widetilde{\mathcal{X}}$ whose i^{th} layer, for $i = 1,\ldots,n$ is modeled as

$$\widetilde{\mathcal{X}}(:,:,i) = [\sum_{j=1}^{k} \mathbf{v}(i,j)A_{(j)} + \mathcal{Z}(:,:,i)] \odot \mathcal{M}_p(:,:,i), \quad (2)$$

where $\mathbf{v}, \{A_{(j)} : j \in \{1,\ldots,k\}\}, k$ are defined as in (1) and $\odot$ denotes the Hadamard or element-wise product. For $x =$

$1, \ldots, m_1$, $y = 1, \ldots m_2$ and $i = 1, \ldots n$, the tensor $\mathcal{M}$ in (2) is defined as

$$\mathcal{M}_p(x, y, i) \quad = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } (1-p), \end{cases}$$

where $p \in (0, 1]$ is the probability of observing a given entry in the signal-plus-noise tensor. Note that when $p = 1$, (2) is equivalent to (1).

III. The Higher Order SVD

If $\mathcal{X} \in \mathbb{R}^{m_1 \times m_2 \times n}$ is a 3-D tensor, then its Higher Order SVD is given by [17, page 95]

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3,$$

where $\mathcal{S} \in \mathbb{R}^{m_1 \times m_2 \times n}$ is the so-called *core tensor* which satisfies the *super orthogonality* property, *i.e.*, $\langle \mathcal{S}(i, :, :), \mathcal{S}(j, :, :) \rangle = \langle \mathcal{S}(:, i, :), \mathcal{S}(:, j, :) \rangle = \langle \mathcal{S}(:, :, i), \mathcal{S}(:, :, j) \rangle = 0$ and $\mathbf{U}_1 \in \mathbb{R}^{m_1 \times m_1}, \mathbf{U}_2 \in \mathbb{R}^{m_2 \times m_2}, \mathbf{U}_3 \in \mathbb{R}^{n \times n}$ are orthogonal matrices which are obtained as

$$\mathbf{U}_d = \text{ LSVD}(\text{unfold}_d(\mathcal{X})). \tag{3}$$

In (3), LSVD(.) returns the left singular vectors of the argument. If $\mathcal{X} \in \mathbb{R}^{m_1 \times m_2 \times n}$, then the unfold operator is defined as [17, page 93-94]

$$\begin{aligned} \text{unfold}_1(\mathcal{X}) &:= [\mathcal{X}(:, 1, :) \ \ldots \ \mathcal{X}(:, m_2, :)] \in \mathbb{R}^{n \times m_1 m_2}, \\ \text{unfold}_2(\mathcal{X}) &:= [\mathcal{X}(:, :, 1) \ \ldots \ \mathcal{X}(:, :, n)] \in \mathbb{R}^{m_2 \times m_1 n}, \\ \text{unfold}_3(\mathcal{X}) &:= \left[\mathcal{X}(1, :, :)^T \ \ldots \ \mathcal{X}(m_1, :, :)^T\right] \in \mathbb{R}^{m_1 \times n m_2}. \end{aligned}$$

The d-mode vectors $\mathcal{X}$ are the column vectors of $\text{unfold}_d(\mathcal{X})$. For a vector $v \in \mathbb{R}^{m_1 m_2 \times 1}$ the operation fold is defined as

$$\text{fold}(v, [m_1, m_2]) := \begin{bmatrix} v(1) & \ldots & v(m_2) \\ \vdots & \vdots & \vdots \\ v((m_1 - 1)m_2 + 1) & \ldots & v(m_1 m_2) \end{bmatrix}.$$

The unfold operation should be viewed as the inverse operation of fold operation. It can be shown that if the *signal tensor* $\mathcal{L}$ has HOSVD given by

$$\mathcal{L} \ = \ \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3,$$

then the matrix $A_{(j)}$ in (1) can be expressed as

$$A_{(j)} = \mathcal{S}(:, :, j) \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2.$$

These $A_{(j)}$ 'images' are the primary HOSVD related objects that appear in detection and classification algorithms that utilize the HOSVD of tensor datacubes; see for e.g. [18, eq. no. (16) page 996] for its appearance in the context of handwritten digit recognition. Thus, it is of interest to quantify the accuracy of the estimates of $A_{(j)}$ formed from the signal-plus-noise tensor $\widetilde{\mathcal{X}}$. To that end, let the HOSVD of the tensor $\widetilde{\mathcal{X}}$ be given by

$$\widetilde{\mathcal{X}} = \widetilde{\mathcal{S}} \times_1 \widetilde{\mathbf{U}}_1 \times_2 \widetilde{\mathbf{U}}_2 \times_3 \widetilde{\mathbf{U}}_{\mathbf{3}}.$$

Then an estimate of $A_{(j)}$, which we denote by $\widetilde{A}_{(j)}$ can be computed from $\widetilde{\mathcal{X}}$ as

$$\widetilde{A}_{(j)} = \widetilde{\mathcal{S}}(:, :, j) \times_1 \widetilde{\mathbf{U}}_1 \times_2 \widetilde{\mathbf{U}}_2. \tag{4}$$

In this paper, we are interested in characterizing how well $A_{(j)}$ is approximated by its estimate $\widetilde{A}_{(j)}$. We shall quantify this using their cosine similarity, which is given by

$$\left\langle \frac{\widetilde{A}_{(j)}}{\|\widetilde{A}_{(j)}\|_F}, \frac{A_{(j)}}{\|A_{(j)}\|_F} \right\rangle = \text{trace}\left(\frac{\widetilde{A}_{(j)}{}^T A_{(j)}}{\|\widetilde{A}_{(j)}\|_F \|A_{(j)}\|_F} \right). \tag{5}$$

In what follows we will theoretically characterize when we can expect the images to be correlated and when they will be uncorrelated.

IV. Main Results

We now state our main results next. In what follows $\xrightarrow{a.s.}$ denotes almost sure convergence.

Theorem 1. *For the setup in (1), let $\theta_1 > \theta_2 > \cdots > \theta_k$. Then as $n, m_1, m_2 \to \infty$ with $m_1 m_2 / n \to c$, we have that*

$$\left\langle \frac{\widetilde{A}_{(j)}}{\|\widetilde{A}_{(j)}\|_F}, \frac{A_{(j)}}{\|A_{(j)}\|_F} \right\rangle^2 \xrightarrow{a.s.} \begin{cases} 1 - \dfrac{c(1+\theta_j^2)}{\theta_j^2(c+\theta_j^2)} & \text{if } \theta_j > c^{\frac{1}{4}}, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\|\widetilde{A}_{(j)}\|_F \xrightarrow{a.s.} \begin{cases} \sqrt{\dfrac{(1+\theta_j^2)(c+\theta_j^2)}{\theta_j^2}} & \text{if } \theta_j > c^{\frac{1}{4}}, \\ 1 + \sqrt{c} & \text{otherwise.} \end{cases}$$

This result implies that the estimate $\widetilde{A}_{(j)}$ is asymptotically localized around the latent $A_{(j)}$ and that this estimate is generically biased and hence inconsistent. An additional insight from Theorem 1 is that the cosine similarity undergoes a *phase transition*, *i.e.*, below a certain critical SNR, the estimate $\widetilde{A}_{(j)}$ is orthogonal to $A_{(j)}$. We now provide an analogous characterization for the missing data setting.

Theorem 2. *In (2), let $\theta_1 > \theta_2 > \cdots > \theta_k$ and assume that $A_{(j)}$ and $\mathbf{v}$ satisfy the low coherence condition, i.e., there exist non-negative constants $\eta_A, \eta_v, C_A, C_{\mathbf{v}}$ independent of n, such that for each $j = 1, \ldots, k$, we have that*

$$\max_j \|A_{(j)}\|_\infty \quad \leq \frac{\eta_A \log^{C_A}(m_1 m_2)}{\sqrt{m_1 m_2}},$$

and

$$\max_j \|\mathbf{v}(:, j)\|_\infty \quad \leq \frac{\eta_v \log^{C_{\mathbf{v}}} n}{\sqrt{n}}.$$

Then for $p \in (0, 1)$, as $n, m_1, m_2 \to \infty$ with $m_1 m_2 / n \to c$, we have that

$$\left\langle \frac{\widetilde{A}_{(j)}}{\|\widetilde{A}_{(j)}\|_F}, \frac{A_{(j)}}{\|A_{(j)}\|_F} \right\rangle^2 \xrightarrow{a.s.} \begin{cases} 1 - \dfrac{c(1+p\,\theta_j^2)}{p\,\theta_j^2(c+p\,\theta_j^2)} & \text{if } \theta_j > \dfrac{c^{\frac{1}{4}}}{\sqrt{p}}, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\|\widetilde{A}_{(j)}\|_F \xrightarrow{a.s.} \begin{cases} \sqrt{\dfrac{(1+p\,\theta_j^2)(c+p\,\theta_j^2)}{\theta_j^2}} & \text{if } \theta_j > \dfrac{c^{\frac{1}{4}}}{\sqrt{p}}, \\ \sqrt{p}\,(1+\sqrt{c}) & \text{otherwise.} \end{cases}$$

Theorem 2 shows that there is a phase transition induced by randomly missing data. The key difference between Theorem 1 and Theorem 2 is that the SNR parameter θ_i is scaled by $\sqrt{p}$ in Theorem 2. In other words, the critical SNR for the phase transition has increased by a factor of $1/\sqrt{p}$. An inspection of Theorem 2 reveals that for fixed θ_j, there is a certain critical value $p_{\text{crit}} = \sqrt{c}/\theta_j^2$ such that if $p < p_{\text{crit}}$, the estimate $\widetilde{A}_{(j)}$ is asymptotically orthogonal to $A_{(j)}$. Above this critical threshold the estimates are correlated. The low coherence assumption echoes the results in [16] and the matrix completion literature.

V. Sketch of the Proof

The results follow from an extension of the recent results from random matrix theory [19], [20]. To that end, we first note that the underlying matrix $A_{(j)}$ and its estimate $\widetilde{A}_{(j)}$ can be written in terms of the matrix unfolding of the signal and corrupted tensor along the third dimension. Let the SVD of the unfolded signal tensor $\mathcal{L}$ be given by

$$\text{unfold}_3\left(\mathcal{L}\right) = \mathbf{U}_3\mathbf{\Theta}_3\mathbf{V}_3^H, \tag{6}$$

where, $\mathbf{U}_3, \mathbf{V}_3$ are, respectively, the left and right singular vector matrices and $\mathbf{\Theta}_3$ is the diagonal matrix of the singular values of the unfolded signal tensor. Analogously, let the SVD of the unfolded signal-plus-noise tensor $\widetilde{\mathcal{X}}$ be given by

$$\text{unfold}_3\left(\widetilde{X}\right) = \widetilde{\mathbf{U}}_3\widetilde{\mathbf{\Theta}}_3\widetilde{\mathbf{V}}_3^H, \tag{7}$$

where $\widetilde{\mathbf{U}}_3, \widetilde{\mathbf{V}}_3$ are, respectively, the left and right singular vector matrices and $\widetilde{\mathbf{\Theta}}_3$ is the diagonal matrix of singular values of the unfolded signal-plus-noise tensor.

The key observation that is a straightforward consequence of tensor algebra is that $A_{(j)}$ can be expressed as

$$\frac{A_{(j)}}{\|A_{(j)}\|_F} = \text{fold}(\mathbf{V}_3(:,j),[m_1,m_2]),$$

with $\|A_{(j)}\|_F = \Theta_3(j,j)$ while $\widetilde{A}_{(j)}$ can be expressed as

$$\frac{\widetilde{A}_{(j)}}{\|\widetilde{A}_{(j)}\|_F} = \text{fold}(\widetilde{\mathbf{V}}_3(:,j),[m_1,m_2]),$$

with $\|\widetilde{A}_{(j)}\|_F = \widetilde{\Theta}_3(j,j)$. Consequently, the cosine similarity between the $A_{(j)}$ and its estimate, $\widetilde{A}_{(j)}$, is given by

$$\left\langle \frac{\widetilde{A}_{(j)}}{\|\widetilde{A}_{(j)}\|}, \frac{A_{(j)}}{\|A_{(j)}\|} \right\rangle = \langle \mathbf{V}_3(:,j), \widetilde{\mathbf{V}}_3(:,j)\rangle.$$

From (6) and (7), we have that,

$$\widetilde{\mathbf{U}}_3\widetilde{\mathbf{\Theta}}_3\widetilde{\mathbf{V}}_3^H = \underbrace{\text{unfold}_3\left(\mathcal{L}\right)}_{=\mathbf{U}_3\mathbf{\Theta}_3\mathbf{V}_3^H} + \text{unfold}_3(\mathcal{Z}),$$

where $\text{unfold}_3(\mathcal{Z})$ is a noise-only matrix with i.i.d. $\mathcal{N}(0,1/n)$ entries and the matrix $\mathbf{U}_3\mathbf{\Theta}_3\mathbf{V}_3^H$ is a rank- k matrix. Thus a direct application of the result in [19, Section 3.1], gives us

$$\left\langle \frac{\widetilde{A}_{(j)}}{\|\widetilde{A}_{(j)}\|_F}, \frac{A_{(j)}}{\|A_{(j)}\|_F} \right\rangle^2 \xrightarrow{a.s.} \begin{cases} 1 - \dfrac{c(1+\theta_j^2)}{\theta_j^2(c+\theta_j^2)} & \text{if } \theta_j > c^{\frac{1}{4}}, \\ 0 & \text{otherwise,} \end{cases}$$

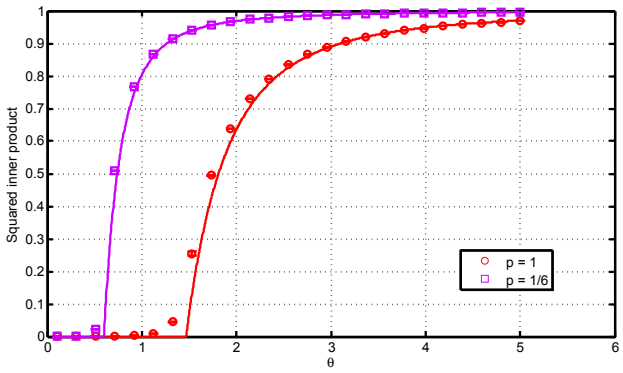


Fig. 1. The cosine similarity as a function of θ for different values of p for the model in (2). Here $k = 1$, $m_1 = m_2 = 50$ and $n = 2500$. The empirical averages were computed over 1000 trials. The solid line is the theoretical prediction from Theorem 2.

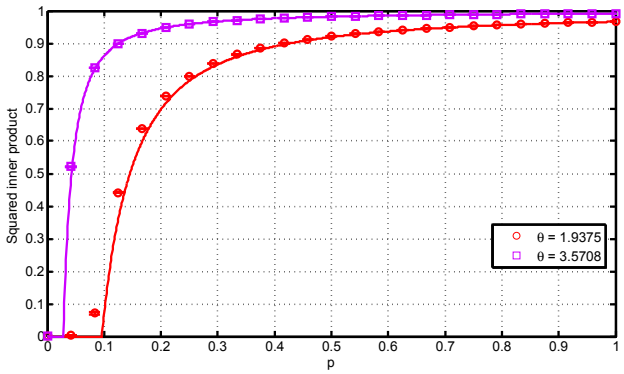


Fig. 2. The cosine similarity as a function of p for different values of θ for the model in (2). Here $k = 1$, $m_1 = m_2 = 50$ and $n = 2500$. The empirical averages were computed over 1000 trials. The solid line is the theoretical prediction from Theorem 2.

as stated in Theorem 1. An application of the same result gives us the stated result for $\|\widetilde{A}_{(j)}\|_F$ in Theorem 1.

For the missing data setting in Theorem 2, the only difference is the presence of the masking matrix $\text{unfold}_3\left(\mathcal{M}_p\right)$ on the right-hand side of (2). Consequently, we have that

$$\widetilde{\mathbf{U}}_3\widetilde{\mathbf{\Theta}}_3\widetilde{\mathbf{V}}_3^H = \left[\mathbf{U}_3\mathbf{\Theta}_3\mathbf{V}_3^H + \text{unfold}_3(\mathcal{Z})\right] \odot \text{unfold}_3\left(\mathcal{M}_p\right).$$

Via the same argument as the preceding Theorem, we can apply the results of [20, Theorem 2.4]. This gives us the results stated in Theorem 2.

VI. Simulation Results

We now validate our theoretical predictions using numerical simulations. For the model in (2), we set $m_1 = 25, m_2 = 40$, $n = 10000$ and generated a random (incoherent) 2-D matrix, $A_{(1)}$, and constructed a 3-D tensor by taking an outer product of this 2-D matrix with a random isotropic vector $\mathbf{v}$ with $\|\mathbf{v}\| = \theta$. This signal tensor was then corrupted with additive Gaussian noise and masked with a binary tensor as in (2). Figure 1 shows the agreement between the prediction of the cosine similarity given by Theorem 2 and the empirical average for the same, computed over 1000 Monte-Carlo trials.

Figure 2 shows the agreement between the theoretical prediction in Theorem 2 and empirical simulation for the model in (2). Figure 3 shows a heatmap of the empirically averaged cosine similarity on a logscale as a function of θ and p. The heatmap clearly illustrates the phase transition and confirms the accuracy of the threshold (shown with the solid white line) predicted in Theorem 2.

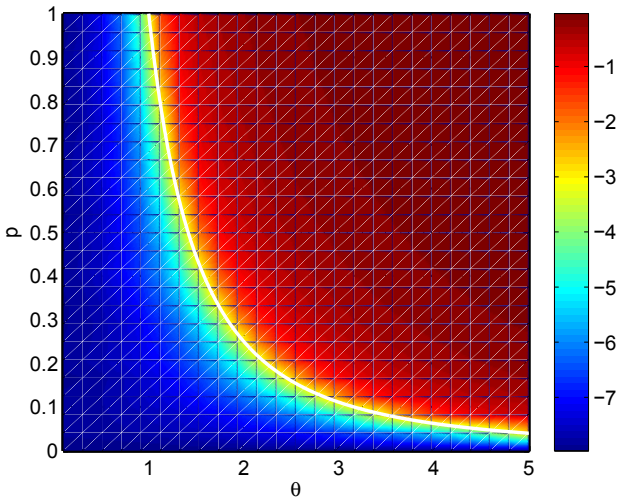
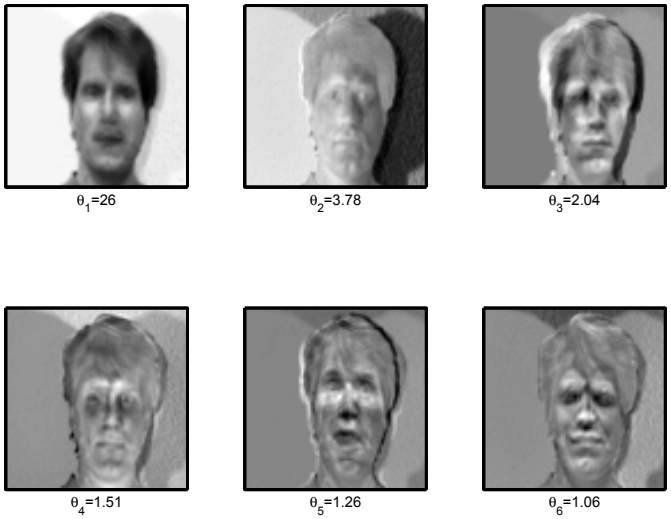
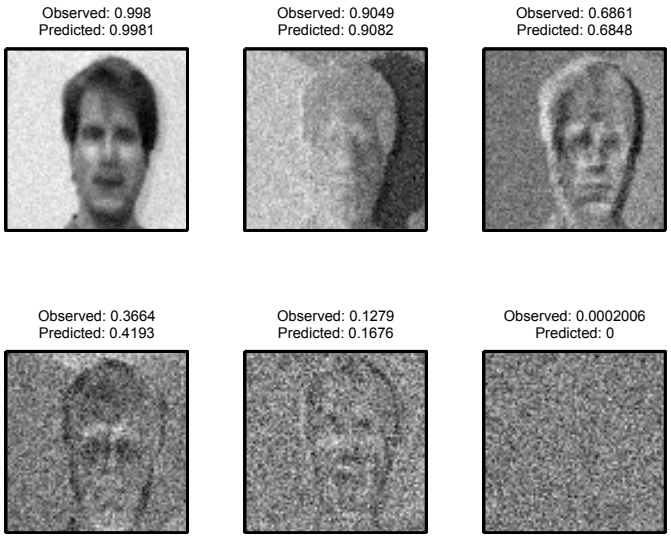


Fig. 3. Empirical heatmap of the logarithm of the cosine similarity, given by (5), averaged over 1000 trials, for the model in (2) with $k = 1$, $m_1 = 25, m_2 = 40$ and $n = 10000$. The deep blue region is the regime where the estimates are nearly orthogonal to the latent signal matrix. The solid white line denotes the phase transition threshold $\theta_{\text{crit}} = c^{1/4}/\sqrt{p}$ as predicted by Theorem 2.

Finally, we present a simulation that shows how the results can provide a principled justification for the empirical observation that the HOSVD is robust to missing data [1]. To illustrate its robustness, we compare the ground-truth $A_{(j)}$'s (see Figure 4(a)) with the $\widetilde{A}_{(j)}$'s (see Figure 4(b)) for images from the *Yale Faces* dataset that were corrupted samples as in (2) with $p = 0.75$ and an associated $\theta = [26, 3.78, 2.04, 1.50, 1.26, 1.06, 0.96, 0.69, 0.56, 0.53, 0.42]$; here $k = 11$ in (1). The shows that there is a strong correlation between the similarity of an estimated representative image $\widetilde{A}_{(j)}$ with the original representative image $A_{(j)}$, and the visual contrast of the estimated representative images. The higher the cosine similarity, the more features are preserved in $\widetilde{A}_{(j)}$. When the cosine similarity is low (for $j = 6$), then the images are uncorrelated as predicted by Theorem 2 and utilizing them in an inferential scheme will result in a degradation in performance.



(a) The ground truth images $A_{(j)}$ for $j = 1, \ldots, 6$.



(b) Estimated images $\widetilde{A}_{(j)}$ for $j = 1, \ldots, 6$.

Fig. 4. The HOSVD is robust to missing data and noise. The ground truth images in (a) were corrupted as in (2) with $p = 0.75$. The images estimated using (4) are shown in (b). The cosine similarity, between the ground truth and the estimated images was computed using (5) and compared with the predictions from Section IV.

VII. Conclusion

We theoretically analyzed the performance of the HOSVD and showed that when the SNR is strong enough then the HOSVD produces estimates of the signal matrix that are correlated with the latent signal matrix. When the SNR is below a critical threshold then the estimates are asymptotically uncorrelated. We analyzed the performance with missing data and showed the presence of a similar critical threshold for the probability of observing an entry, below which the estimates were uncorrelated. We corroborated the theoretical predictions with empirical simulations. A natural next step is to analyze the performance of other tensor decomposition schemes such as the PARAFAC/CANDECOMP [1]. This can facilitate a comparison of their performance relative to the fundamental limits of the HOSVD uncovered in this paper and provide a better understanding of which technique(s) best uncover weak latent structure in the presence of missing data from noisy datacubes.

Acknowledgment

This work was supported by an ONR Young Investigator Award N000141110660, an AFOSR Young Investigator Award

FA9550-12-1-0266, a NSF award CCF-1116115 and a ARO MURI grant W911NF-11-1-0391.

REFERENCES

[1] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[2] A. Cichocki, Y. Washizawa, T. M. Rutkowski, H. Bakardjian, A. H. Phan, S. Choi, H. Lee, Q. Zhao, L. Zhang, and Y. Li, "Noninvasive bcis: Multiway signal-processing array decompositions." *IEEE Computer*, vol. 41, no. 10, pp. 34–42, 2008.

[3] G. Bergqvist and E. G. Larsson, "The higher-order singular value decomposition: Theory and an application [lecture notes]," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 151–154, 2010.

[4] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. on Signal Processing*, vol. 48, no. 8, pp. 2377–2388, 2000.

[5] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *Computer Vision - ECCV 2002*, pp. 447–460, 2002.

[6] C. Tian, G. Fan, X. Gao, and Q. Tian, "Multiview face recognition: from tensorface to v-tensorface and k-tensorface," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 320–333, 2012.

[7] D. Terzopoulos, Y. Lee, and M. A. O. Vasilescu, "Model-based and image-based methods for facial image synthesis, analysis and recognition,", pp. 3–8, *Proc., Sixth IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2004.

[8] T. G. Kolda and J. Sun, "Scalable tensor decompositions for multi-aspect data mining," *Eighth IEEE International Conference on Data Mining*, pp. 363–372, 2008,

[9] A. Karami, M. Yazdi, and G. Mercier, "Compression of hyperspectral images using discerete wavelet transform and tucker decomposition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 444–450, 2012.

[10] M. Signoretto, R. Van de Plas, B. De Moor, and J. A. Suykens, "Tensor versus matrix completion: a comparison with application to spectral data," *IEEE Signal Processing Letters*, vol. 18, no. 7, pp. 403–406, 2011.

[11] M. Weis, F. Romer, M. Haardt, D. Jannek, and P. Husar, "Multi-dimensional space-time-frequency component analysis of event related eeg data using closed-form parafac," *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing,* pp. 349–352, 2009.

[12] Y. Washizawa, H. Higashi, T. Rutkowski, T. Tanaka, and A. Cichocki, "Tensor based simultaneous feature extraction and sample weighting for eeg classification," in *Neural Information Processing. Models and Applications*, pp. 26–33, 2010.

[13] M. Weis, D. Jannek, T. Guenther, P. Husar, F. Roemer, and M. Haardt, "Temporally resolved multi-way component analysis of dynamic sources in event-related eeg data using parafac2," in *Proceedings of the 18th European Signal Processsing Conference (EUSIPCO-2010), Aalborg, Denmark*, 2010.

[14] S. L. Freire and T. J. Ulrych, "Application of singular value decomposition to vertical seismic profiling," *Geophysics*, vol. 53, no. 6, pp. 778–785, 1988.

[15] J.-F. Cardoso, "Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem," in *Proc. of IEEE Conf. on Acoustics, Speech, and Signal Processing,*, pp. 2655–2658, 1990.

[16] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, p. 11, 2011.

[17] L. Eldén, *Matrix methods in data mining and pattern recognition*, vol. 4, SIAM, 2007.

[18] B. Savas and L. Eldén, "Handwritten digit classification using higher order singular value decomposition," *Pattern recognition*, vol. 40, no. 3, pp. 993–1003, 2007.

[19] F. Benaych-Georges and R. R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices," *Journal of Multivariate Analysis*, vol. 111, pp. 120–135, 2012.

[20] R. Nadakuditi, "Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 1–17, May 2014.