Exact Tensor Completion Using t-SVD

Zemin Zhang, *Student Member, IEEE*, and Shuchin Aeron, *Senior Member, IEEE*

***Abstract*—In this paper, we focus on the problem of completion of multidimensional arrays (also referred to as tensors), in particular three-dimensional (3-D) arrays, from limited sampling. Our approach is based on a recently proposed tensor algebraic framework where 3-D tensors are treated as linear operators over the set of 2-D tensors. In this framework, one can obtain a factorization for 3-D data, referred to as the tensor singular value decomposition (t-SVD), which is similar to the SVD for matrices. t-SVD results in a notion of rank referred to as the tubal-rank. Using this approach we consider the problem of sampling and recovery of 3-D arrays with low tubal-rank. We show that by solving a convex optimization problem, which minimizes a convex surrogate to the tubal-rank, one can guarantee exact recovery with high probability as long as number of samples is of the order $O(rnk\log(nk))$, given a tensor of size $n \times n \times k$ with tubal-rank r. The conditions under which this result holds are similar to the incoherence conditions for low-rank matrix completion under random sampling. The difference is that we define incoherence under the algebraic setup of t-SVD, which is different from the standard matrix incoherence conditions. We also compare the numerical performance of the proposed algorithm with some state-of-the-art approaches on real-world datasets.**

***Index Terms*—Tensor completion, sampling and recovery, convex optimization.**

I. Introduction

RECOVERY of multidimensional array of numbers or *tensors*[1] under limited number of measurements is an important problem, which arises in a variety of applications, such as recommendation systems [2], dimensionality reduction [3], multi-class learning [4], data mining [5], [6], and computer vision [7], [8].

The strategies and performance bounds for sampling and recovery of tensors rest heavily on the framework used to reveal a low complexity algebraic structure in the data, namely a *low-rank* decomposition. For example, for matrix data (a 2-dimensional tensor) when treated as a linear operator over a vector space, one defines the rank of the matrix via its minimal decomposition into a sum of rank-1 matrices. This is well-known to be obtained via the Singular Value Decomposition (SVD). In this case, it has been shown that as long as the left and right singular vectors are incoherent with the standard basis, a random sampling strategy with sampling complexity in proportion to the complexity of the decomposition is sufficient for recovery by solving a convex optimization problem, namely minimizing the nuclear norm of the matrix [9], [10].

For N-D tensors with $N \geq 3$ (**Note**: N is also referred to as the order of the tensor, and we will often refer to an N-D tensor as an order-N tensor), there are several ways to define algebraic complexity using the framework of classical multilinear algebra, where tensors are treated as multilinear operators over tensor product or outer product of vector spaces [1]. In this framework, decomposition of a tensor as sum of rank-1 *outer products* is referred to as CANDECOMP/PARAFAC (CP) [11] factorization and the minimal number of such factors required is referred to as the CP rank. However there are known computational and ill-posedness issues with CP [12]. Other kinds of decompositions, such as Tucker, Hierarchical-Tucker (H-Tucker) and Tensor Train (TT) [13], are also shown to reveal the algebraic structure in the data with the notion of rank extended to *multi-rank*, expressed as a vector of ranks of the factors in the contracted representation using matrix product states. In this context, to the best of our knowledge, provably optimal approaches based on Tucker decomposition work by *matricizing* the tensor in various ways, subsequently employing the theory and methods for matrix completion, see for example [14], [15].

This paper considers sampling and recovery for 3-D tensors using the algebraic framework proposed in [16]–[19]. In this framework, which we outline in Section II, 3-D tensors are treated as linear operators over 2-D tensors and one obtains an SVD-like factorization referred to as the tensor-SVD (t-SVD). Using this factorization, one can define a notion of rank, referred to as the tubal-rank. This algebraic framework is essentially based on a group theoretic approach where the multidimensional structure is unraveled by constructing group-rings along the tensor fibers [20]. This approach has recently been extended in [21] to construct a Banach algebra along tensor fibers. In this paper, we restrict ourselves to group rings constructed out of cyclic groups and also omit consideration of the extensions carried out in [21]. Nevertheless, the results presented here can be generalized to these settings.

It is important to note that the t-SVD algebraic framework is different from the classic multilinear algebraic framework for tensor decompositions [1]. Therefore, the notion of tensor rank using the t-SVD, namely the tubal-rank, differs from the CP rank and the Tucker rank. Hence, bounds and conditions for tensor completion for low CP rank and low Tucker rank tensors are not *directly* comparable to results in this paper.

The t-SVD has been recently exploited in [8] for the problem of 3-D tensor recovery from limited sampling with applications

Manuscript received January 8, 2016; revised July 13, 2016 and November 1, 2016; accepted November 3, 2016. Date of publication December 14, 2016; date of current version January 20, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gesualdo Scutari. This work was supported by the National Science Foundation under Grant CCF:1319653.

The authors are with the Department of Electrical and Computer Engineering, Tufts University, Medford, MA 02155 USA (e-mail: zemin.zhang@tufts.edu; shuchin@ece.tufts.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2016.2639466

[1]This terminology comes from representation of multilinear functionals on the outer product of finite dimensional vector spaces as an indexed array of numbers, for example see [1].

TABLE I
A SUMMARY OF EXISTING TENSOR COMPLETION METHODS

Format	**Sampling Method**	**Samples needed for exact recovery (3rd-order tensor of size $n \times n \times n$)**	**Incoherent condition**
CP[23]	Gaussian measurements	$O(rn^2)$ for CP rank r	N/A
CP[24]	Random sampling	$O(n^{3/2} r^5 \log^4(n))$ for CP rank r on symmetric tensors	Incoherence condition of symmetric tensors with orthogonal decomposition
Tucker[23]	Gaussian measurements	$O(rn^2)$ for Tucker rank (r, r, r)	N/A
Tucker[14]	Random sampling	$O(rn^2 \log^2(n))$ for Tucker rank (r, r, r)	Matrix incoherence condition on all mode-n unfoldings
CP[25]	Adaptive Sampling	$O(nr \log(r))$ for CP rank r	Standard incoherence condition with orthogonal decomposition
t-SVD (this paper)	Random sampling	$O(rn^2 \log(n))$ for Tensor tubal-rank r	Tensor incoherence condition

to computer vision. In the present paper, we derive theoretical performance bounds for the tensor recovery algorithm proposed in [8]. In this context, our work is greatly inspired by [10], [22], in which the main tool, namely the Non-commutative Bernstein Inequality (NBI), is also helpful in deriving our results. We prove that with high probability one can exactly recover a tensor of size $n_1 \times n_2 \times n_3$ with tubal-rank r (as derived from the t-SVD, see Section II), by solving a convex optimization problem, given $\mathcal{O}(rn_1 n_3 \log((n_1 + n_2) n_3))$ samples when certain tensor incoherence conditions are satisfied. The notions of tensor incoherence and results are novel, and we show that, while related, they are *not* directly implied by the results in matrix completion using the standard matrix incoherence conditions.

In order to put our work into perspective and highlight our contributions, we now go over related work on tensor completion using different tensor factorizations and contrast our findings with existing literature.

A. Related Work

Apart from the t-SVD, there are two major types of low-rank tensor completion methods considered in the literature: methods that are based on the CP decomposition, and those that are based on the Tucker decomposition. The sampling methods include random downsampling, Gaussian measurements and adaptive sampling. We summarize these results in **Table** I. Below we will provide details for each of these methods.

1) Tensor Completion Based on CP Decomposition: The CP decomposition of an order-N tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_N}$ is given by,

$$\mathcal{X} = \sum_{\ell=1}^{r} \mathbf{x}_\ell^{(1)} \circ \mathbf{x}_\ell^{(2)} \circ \cdots \circ \mathbf{x}_\ell^{(N)}, \;\; \mathbf{x}_\ell^{(i)} \in \mathbb{R}^{n_i}, i = 1, 2, \ldots, N, \tag{1}$$

where $\circ$ denotes the outer product [11]. The smallest r such that Equation (1) holds is called the CP rank of $\mathcal{X}$.

Suppose we sample $\mathcal{X}$ based on an index set Ω. Let P_Ω be the orthogonal projection onto Ω. Then in [26] the authors propose to complete the tensor by solving the following optimization problem,

$$\min \|P_\Omega \left(\mathcal{A} - \sum_{\ell=1}^{r} \mathbf{x}_\ell^{(1)} \circ \mathbf{x}_\ell^{(2)} \circ \ldots \circ \mathbf{x}_\ell^{(N)} \right) \|_2^2 + \lambda \sum_{\ell=1}^{r} \sum_{i=1}^{N} \|\mathbf{x}_\ell^{(i)}\|_2^2 \ , \tag{2}$$

where $\lambda \geq 0$ is the regularization parameter. However, this approach has several drawbacks. The optimization problem is non-convex and hence only local minima can be guaranteed. Further, for practical problems it is often computationally difficult to determine the CP rank or the best low rank CP approximation of a tensor data beforehand. Recently in [24] it was shown that one can provably recover an $n \times n \times n$ *symmetric* tensor with CP rank r from $O(n^{3/2} r^5 \log^4 n)$ randomly sampled entries under standard incoherence conditions on the factors.[2]

2) Tensor Completion Based on Tucker Decomposition: In [27] tensor completion based on minimizing a convex surrogate to the tensor n-rank is proposed. Tensor n-rank is the sum of the ranks of matrices obtained by the matricizations of the tensor, i.e. it is the sum of the Tucker ranks. Specifically one solves for,

$$\min_{\mathcal{X}} \sum_{i=1}^{N} \alpha_i \|\mathcal{X}_{(i)}\|_* \quad \text{subject to } P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}) \ , \tag{3}$$

where $\mathcal{X}_{(i)}$ denotes the mode-i matricization of $\mathcal{X}$ [11], α_i are pre-specified positive constants satisfying $\sum_{i=1}^{N} \alpha_i = 1$, and $\| \cdot \|_*$ denotes the matrix nuclear norm. However, no theoretical guarantees for recovery are provided and it is not clear how to optimally choose the weights α_i's. Normally one ends up choosing the best matricization that is determined empirically, which turns it into a matrix completion problem. In [28] a tighter convex relaxation for the tensor n-rank is proposed, but again no provable recovery bounds are provided.

In [14] the authors solve the following convex problem,

$$\min_{\mathcal{X}, \mathcal{E}} \sum_{i=1}^{N} \alpha_i \|\mathcal{X}_{(i)}\|_* + \|\mathcal{E}\|_1 + \frac{\tau}{2}\|\mathcal{X}\|_F^2 + \frac{\tau}{2}\|\mathcal{E}\|_F^2 \quad \text{s.t. } P_\Omega(\mathcal{X} + \mathcal{E}) = \mathcal{B}, \tag{4}$$

for some specific choices for α_i and τ. This can be viewed as a combination of the matrix completion and the matrix Robust Principal Component Analysis (RPCA) when extended to the case of tensors. It is shown that if the tensor satisfies the matrix incoherence conditions under all its matricizations, then solving for the above optimization problem leads to accurate recovery

[2] A tensor $\mathcal{X} \in \mathbb{R}^{n \times n \times n}$ is called symmetric in the CP format if its CP decomposition has the format $\mathcal{X} = \sum_{\ell=1}^{r} \sigma_\ell (\mathbf{u}_\ell \circ \mathbf{u}_\ell \circ \mathbf{u}_\ell)$, where $\mathbf{u}_\ell \in \mathbb{R}^n$ with $\|\mathbf{u}_\ell\| = 1$.

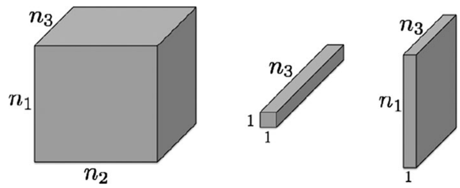


Fig. 1. General third-order tensor, tensor tube and tensor column.

if the number of samples exceeds that required for completion under each matricization.

3) Tensor Completion Under Gaussian Measurements: Instead of random sampling, in [23] a different method for low CP or Tucker rank tensor completion under Gaussian measurements is proposed. The main idea is to *reshape* the tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \ldots \times n_N}$ into a *square* matrix $\mathcal{X}_{[j]} \in \mathbb{R}^{n_1 \ldots n_j \times n_{j+1} \cdots N}$ and apply matrix completion on it. If $\mathcal{X}$ is a low-rank tensor (in either CP or Tucker sense), $\mathcal{X}_{[j]}$ will be a low-rank matrix. It is shown that if $\mathcal{X}_0$ has CP rank r, then $O(rn^{\lceil N/2 \rceil})$ Gaussian samples are sufficient to recover the original tensor. If $\mathcal{X}_0$ has Tucker rank $(r, r, \ldots, r)$, then $O(r^{\lfloor N/2 \rfloor} n^{\lceil N/2 \rceil})$ Gaussian samples are needed.

4) Tensor Completion via Adaptive Sampling: In [25] a tensor completion approach based on adaptive sampling is developed. The key idea is to predict the tensor singular subspace given the sampled sub-tensor, and recursively update the subspace if a newly sampled sub-tensor lies out of it. It is shown that $O(nr^{3/2} \log r)$ adaptively chosen samples are sufficient for exact recovery of an $n \times n \times n$ tensor with CP rank r. This approach extends the matrix completion to the tensor case and yields a tighter bound, requiring only column incoherence conditions.

B. *Organization of the Paper*

This paper is organized as follows. Section II introduces the notations and outlines the algebraic framework for tensor factorization that is used in this paper. In Section III we present the main result on tensor completion and derive provable performance bounds for tensor completion under low tubal-rank as defined through the t-SVD. We then provide the full proof in Section IV and report empirical and numerical results in Section V. Finally we outline the implications of the work and future research in Section VI.

II. Notations and Preliminaries

Matrices are represented by uppercase boldface letters and vectors by lower case boldface letters. Tensors are represented in bold script font. For instance, a third-order tensor is represented as $\mathcal{A}$, and its $(i, j, k)th$ entry is represented as $\mathcal{A}_{ijk}$. Moreover, a tensor tube of size $1 \times 1 \times n_3$ is denoted as $\mathring{\boldsymbol{a}}$, and a tensor column of size $n_1 \times 1 \times n_3$ is denoted as $\vec{\boldsymbol{b}}$. These are illustrated in Figure 1. Often we will use the following notation to extract the tensor slices/fibers. $\mathcal{A}(:, :, i)$ denotes the i-th frontal slice obtained by varying all but the third index, $\mathcal{A}(i, j, :)$ denotes a tube/fiber oriented into the board obtained by fixing the first two indices and varying the third, and so on. For the frontal slices, we will frequently use a more compact notation $\mathcal{A}^{(i)} \triangleq \mathcal{A}(:, :, i) \in \mathbb{R}^{n_1 \times n_2}$.

$\widehat{\mathcal{A}}$ denotes a third-order tensor obtained by taking the Fourier Transform of all the tubes along the third dimension of $\mathcal{A}$, i.e., for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$, i.e.,

$$vec(\widehat{\mathcal{A}}(i, j, :)) = \mathfrak{F}(vec(\mathcal{A}(i, j, :))), \tag{5}$$

where vec is the vectorization operator that takes the tensor tube and makes it a vector, and $\mathfrak{F}$ stands for the Discrete Fourier Transform (DFT). For compactness, we will use the following notation for the DFT along the 3rd dimension: $\widehat{\mathcal{A}} = \texttt{fft}(\mathcal{A}, [\,], 3)$. In the same fashion, one can also compute $\mathcal{A}$ from $\widehat{\mathcal{A}}$ via $\texttt{ifft}(\widehat{\mathcal{A}}, [\,], 3)$ using the inverse FFT operation along the 3-rd dimension. For vectors $\mathbf{x}$ and $\mathbf{y}$ of length n, $\mathbf{y} = \texttt{fft}(\mathbf{x})$ and $\mathbf{x} = \texttt{ifft}(\mathbf{y})$ are defined as follows.

$$\begin{aligned} \mathbf{y}(k) &= \sum_{j=1}^{n} \mathbf{x}(j) w_n^{(j-1)(k-1)} \\ \mathbf{x}(j) &= \frac{1}{n} \sum_{k=1}^{n} \mathbf{y}(k) w_n^{-(j-1)(k-1)} \ , \end{aligned} \tag{6}$$

where $w_n = e^{(-2\pi i)/n}$ is one of the n roots of unity.

Definition II.1 (Tensor transpose [17]): The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the $n_2 \times n_1 \times n_3$ tensor $\mathcal{A}^\top$ obtained by conjugate transposing each of the frontal slice and then reversing the order of transposed frontal slices 2 through n_3:

$$\begin{aligned} \left(\mathcal{A}^\top\right)^{(1)} &= \left(\mathcal{A}^{(1)}\right)^\top \\ \left(\mathcal{A}^\top\right)^{(i)} &= \left(\mathcal{A}^{(n_3+2-i)}\right)^\top, \ i = 2, \ldots, n_3 \end{aligned}$$

Definition II.2 (t-product [17]): The t-product $\mathcal{A} * \mathcal{B}$ of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ is an $n_1 \times n_4 \times n_3$ tensor whose (i, j)th tube $\mathring{\boldsymbol{c}}_{ij}$ is given by

$$\mathring{\boldsymbol{c}}_{ij} = \mathcal{C}(i, j, :) = \sum_{k=1}^{n_2} \mathcal{A}(i, k, :) * \mathcal{B}(k, j, :) \tag{7}$$

where $*$ denotes the circular convolution between two tubes of same size.

Interpreted in another way, a 3-D tensor of size $n_1 \times n_2 \times n_3$ can be viewed as an $n_1 \times n_2$ matrix of fibers (tubes) oriented along the third dimension. So the t-product of two tensors can be regarded as a matrix-matrix multiplication, except that the multiplication operation between scalars is replaced by circular convolution between the tubes. This allows one to consider 3-D tensors as linear operators over matrices. That is, when $\mathcal{B}$ is an $n_2 \times 1 \times n_3$ tensor (essentially matrix oriented into the paper), $\mathcal{A} * \mathcal{B}$ is an $n_1 \times 1 \times n_3$ tensor. This perspective has been recently used in [29] for problems of unsupervised clustering of 2-D data. For sake of brevity we direct the interested readers to [16], [21].

Definition II.3 (Identity tensor [17]): The identity tensor $\mathcal{I} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ is defined to be a tensor whose first frontal slice $\mathcal{I}^{(1)}$ is the $n_1 \times n_1$ identity matrix and all other frontal slices $\mathcal{I}^{(i)}, i = 2, \ldots, n_3$ are zero.

Definition II.4 (Orthogonal tensor [17]): A tensor $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies

$$\mathcal{Q}^\top * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^\top = \mathcal{I} \tag{8}$$

Algorithm 1: t-SVD for Third Order Tensors.

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

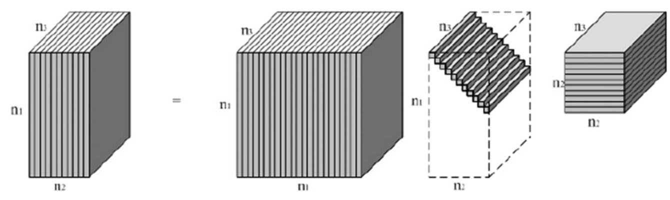


Fig. 2. The t-SVD of an $n_1 \times n_2 \times n_3$ tensor.

Definition II.5 (Block diagonal form of third-order tensor [8]): Let $\overline{\mathcal{A}}$ denote the block-diagonal matrix of the tensor $\widehat{\mathcal{A}}$ in the Fourier domain, i.e.,

$$\overline{\mathcal{A}} \triangleq \text{blockdiag}(\widehat{\mathcal{A}})$$
$$\triangleq \begin{bmatrix} \widehat{\mathcal{A}}^{(1)} & & & \\ & \widehat{\mathcal{A}}^{(2)} & & \\ & & \ddots & \\ & & & \widehat{\mathcal{A}}^{(n_3)} \end{bmatrix} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3} \tag{9}$$

It is easy to verify that the block diagonal matrix of $\mathcal{A}^\top$ is equal to the transpose of the block diagonal matrix of $\mathcal{A}$:

$$\overline{\mathcal{A}^\top} = \overline{\mathcal{A}}^\top \tag{10}$$

Remark II.1: The following fact will be used through out the paper. For any tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, we have

$$\mathcal{A} * \mathcal{B} = \mathcal{C} \Longleftrightarrow \overline{\mathcal{A}\mathcal{B}} = \overline{\mathcal{C}}$$

The t-product allows us to define a tensor Singular Value Decomposition (t-SVD). We need one more definition to state the decomposition.

Definition II.6 (f-diagonal tensor [17]): A tensor $\mathcal{A}$ is called f-diagonal if each frontal slice $\mathcal{A}^{(i)}$ is a diagonal matrix.

Definition II.7 (Tensor Singular Value Decomposition: t-SVD [16], [17]): For $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the t-SVD of $\mathcal{M}$ is given by

$$\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top \tag{11}$$

where $\mathcal{U}$ and $\mathcal{V}$ are orthogonal tensors of size $n_1 \times n_1 \times n_3$ and $n_2 \times n_2 \times n_3$ respectively. $\mathcal{S}$ is a rectangular f-diagonal tensor of size $n_1 \times n_2 \times n_3$, and the entries in $\mathcal{S}$ are called the singular values of $\mathcal{M}$. $*$ denotes the t-product here.

One can obtain this decomposition by computing matrix SVDs in the Fourier domain as shown in Algorithm 1. Figure 2 illustrates the decomposition for the 3-D case.

Based on the t-SVD one can define the following notion of tensor rank.

Definition II.8 (Tensor tubal-rank [8]): The tensor **tubal-rank** r of $\mathcal{A}$ is defined to be the number of non-zero singular tubes of $\mathcal{S}$, where $\mathcal{S}$ comes from the t-SVD of $\mathcal{A}: \mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$. An alternative definition of tubal-rank is that it is the largest rank of all the frontal slices of $\widehat{\mathcal{A}}$ in Fourier domain. If we say a third order tensor $\mathcal{A}$ is of **full tubal-rank**, it means $r = \min\{n_1, n_2\}$.

Remark II.2: It is usually sufficient to compute the reduced version of the t-SVD using the tensor tubal-rank. It's faster and more economical for storage. In details, suppose $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ has tensor tubal-rank r, then the reduced t-SVD of $\mathcal{M}$ is given by

$$\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top \tag{12}$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times r \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times r \times n_3}$ satisfying $\mathcal{U}^\top * \mathcal{U} = \mathcal{I}, \mathcal{V}^\top * \mathcal{V} = \mathcal{I}$. $\mathcal{S}$ is an f-diagonal tensor of size $r \times r \times n_3$. This reduced version of the t-SVD will be used throughout the paper unless otherwise noted.

An important property of the t-SVD is the optimality of the truncated t-SVD for data approximation as stated in the following Lemma.

Lemma II.1 ([17], [18]): Let the t-SVD of $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be given by $\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$ and for $k < \min(n_1, n_2)$ define $\mathcal{M}_k = \sum_{i=1}^{k} \mathcal{U}(:, i, :) * \mathcal{S}(i, i, :) * \mathcal{V}(:, i, :)^T$, Then

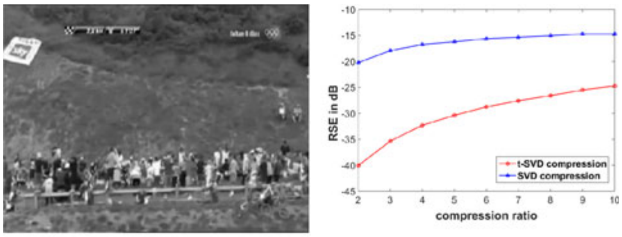
$$\mathcal{M}_k = \arg\min_{\widetilde{\mathcal{M}} \in \mathbb{M}} \|\mathcal{M} - \widetilde{\mathcal{M}}\|_F$$

where $\mathbb{M} = \{\mathcal{C} = \mathcal{X} * \mathcal{Y} | \mathcal{X} \in \mathbb{R}^{n_1 \times k \times n_3}, \mathcal{Y} \in \mathbb{R}^{k \times n_2 \times n_3}\}$ is the set of all tensors with tensor tubal-rank at most k and $\|\cdot\|_F$ is the Frobenius norm defined in Definition II.14.

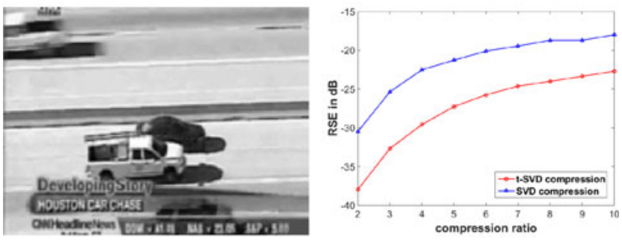
Based on Lemma II.1, one can perform dimensionality reduction and hence tensor compression using the truncated t-SVD. Specifically, given a tensor $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, as in Lemma II.1, we take the rank k approximation $\mathcal{X}_k = \sum_{i=1}^{k} \mathcal{U}(:, i, :)\mathcal{S}(i, i, :)\mathcal{V}(:, i, :)^\top$ as a compression of $\mathcal{X}$. Define the compression ratio as the number of entries used to represent $\mathcal{X}_k$ over the number of entries in $\mathcal{X}$,

$$\text{ratio} = \frac{n_1 n_2 n_3}{n_1 k n_3 + k n_3 + k n_2 n_3} = \frac{n_1 n_2}{k(n_1 + n_2 + 1)}$$

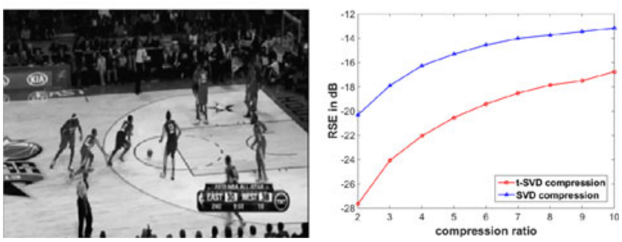
In Figure 3 we provide some video examples, which can be compressed (approximated) well using the truncated t-SVD, compared to vectorizing or flattening the data and using the truncated SVD [8]. Note that we don't compare to other tensor decompositions such as CP or Tucker since it is known that truncated CP or truncated Tucker is not the best low-rank approximation in contrast to the truncated SVD and t-SVD. From the Relative Square Error (RSE) versus the compression ratio plots, we see better performance of the t-SVD over SVD in compression. Here all the videos share a similar feature that the camera is horizontally panning and with linear motion in the video as well. In such videos, a compact representation of one frame to the next frame can be effectively represented as a shift operation, which is captured by a convolution type operation. The t-SVD is based on such an operation along the third



(a) Bicycle video of size $270 \times 480 \times 50$.



(b) Car video of size $136 \times 240 \times 50$.



(c) Basketball video of size $360 \times 640 \times 80$.

Fig. 3. Some videos and their compression performance using the SVD and the t-SVD. The left figures show one frame of each video and the right figures are the compression performance comparisons of the SVD and t-SVD. The relative square error (RSE) is defined in dB as $\text{RSE} = 20\log_{10}(\|\boldsymbol{\mathcal{X}}_{\text{com}} - \boldsymbol{\mathcal{X}}\|_F / \|\boldsymbol{\mathcal{X}}\|_F)$, where $\boldsymbol{\mathcal{X}}$ is the original video and $\boldsymbol{\mathcal{X}}_{\text{com}}$ is the compressed video, and tensor Frobenius norm $\|\cdot\|_F$ is defined in Definition II.14.

dimension, so it results in a much efficient representation. Since efficiency in representation implies efficiency in recovery, as we will show later, such data will have a better performance in completion using the t-SVD as well.

Remark II.3 (Relation to CP decomposition): Suppose a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ has CP rank r and its CP decomposition is given by

$$\boldsymbol{\mathcal{A}} = \sum_{i=1}^{r} \mathbf{a}_i^{(1)} \circ \mathbf{a}_i^{(2)} \circ \mathbf{a}_i^{(3)}$$

where $\mathbf{a}_i^{(k)} \in \mathbb{R}^{n_k}, k = 1, 2, 3$. Then tensor $\widehat{\boldsymbol{\mathcal{A}}}$ which is obtained by taking the FFT along the third dimension of $\boldsymbol{\mathcal{A}}$, has the CP decomposition as follows,

$$\widehat{\boldsymbol{\mathcal{A}}} = \sum_{i=1}^{r} \mathbf{a}_i^{(1)} \circ \mathbf{a}_i^{(2)} \circ \widehat{\mathbf{a}_i^{(3)}}$$

where $\widehat{\mathbf{a}_i^{(3)}} = \texttt{fft}(\mathbf{a}_i^{(3)}), i = 1, 2, \ldots, r$. We can see that $\widehat{\boldsymbol{\mathcal{A}}}$ also has CP rank r, and each frontal slice of $\widehat{\boldsymbol{\mathcal{A}}}$ is the sum of r rank-1 matrices, so the rank of each frontal slice is at most r. It implies that if a tensor is of CP rank r, its tensor tubal-rank is at most r. This means that for a third-order tensor with low CP rank, as we will show later, we can recover it from random samples using the t-SVD structure.

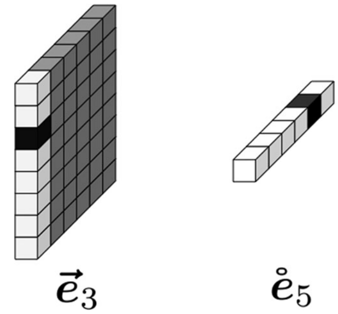


Fig. 4. The column basis $\vec{e}_3$ and tube basis $\mathring{e}_5$. The black cubes are 1, gray and white cubes are 0. The white cubes stand for the potential entries that could be 1.

Definition II.9 (Inverse of tensor [17]): The inverse of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n \times n \times n_3}$ is written as $\boldsymbol{\mathcal{A}}^{-1}$ satisfying

$$\boldsymbol{\mathcal{A}}^{-1} * \boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{A}}^{-1} = \mathfrak{I} \tag{13}$$

where $\mathfrak{I}$ is the **identity tensor** of size $n \times n \times n_3$.

Definition II.10 (Tensor operator): Tensor operators are denoted by Calligraphic letters. Suppose $\mathcal{L} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{n_4 \times n_2 \times n_3}$ is a tensor operator mapping an $n_1 \times n_2 \times n_3$ tensor $\boldsymbol{\mathcal{A}}$ to an $n_4 \times n_2 \times n_3$ tensor $\boldsymbol{\mathcal{B}}$ via the t-product as follows:

$$\boldsymbol{\mathcal{A}} = \mathcal{L}(\boldsymbol{\mathcal{B}}) = \mathcal{L} * \boldsymbol{\mathcal{B}} \tag{14}$$

where $\mathcal{L}$ is an $n_4 \times n_1 \times n_3$ tensor. Note that (14) is equivalent to the following equation, which lies in the Fourier domain:

$$\overline{\boldsymbol{\mathcal{A}}} = \overline{\mathcal{L}}\overline{\boldsymbol{\mathcal{B}}} \tag{15}$$

where $\overline{\boldsymbol{\mathcal{A}}} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3}$, $\overline{\mathcal{L}} \in \mathbb{C}^{n_4 n_3 \times n_1 n_3}$ and $\overline{\boldsymbol{\mathcal{B}}} \in \mathbb{C}^{n_4 n_3 \times n_2 n_3}$ are block diagonal matrices.

Remark II.4: For a tensor operator via t-product defined in (14), we are able to transform it into the equivalent form in Fourier domain (15) for computational efficiency. On the other hand, we can also transform an operator in Fourier domain back to the original domain as needed.

Definition II.11 (Inner product of tensors): If $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ are third-order tensors of same size $n_1 \times n_2 \times n_3$, then the inner product between $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ is defined as the following,

$$\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle = \frac{1}{n_3} \text{trace}(\overline{\boldsymbol{\mathcal{B}}}^\top \overline{\boldsymbol{\mathcal{A}}}) \in \mathbb{R} \tag{16}$$

where $1/n_3$ comes from the normalization constant of the FFT. The reason that this inner product produces a real-valued result comes from the conjugate symmetric property of the FFT.

Definition II.12 (Tensor basis and the corresponding decomposition): We introduce 2 tensor bases here. The first one is called **column basis** $\vec{e}_i$ of size $n_1 \times 1 \times n_3$ with one entry equaling 1 and the rest equaling zero. However, the nonzero entry 1 will only appear at the first frontal slice of $\vec{e}_i$. Naturally its transpose $\vec{e}_i^\top$ is called **row basis**. The other tensor basis is called **tube basis** $\mathring{e}_i$ of size $1 \times 1 \times n_3$ with one entry equaling to 1 and rest equaling to 0. Figure 4 illustrates these 2 bases.

One can obtain a unit tensor $\mathcal{E}$ with the only non-zero entry $\mathcal{E}_{ijk}$ equaling to 1 via

$$\mathcal{E} = \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top \ . \tag{17}$$

Given any third order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have the following decomposition

$$\begin{aligned}\mathcal{X} &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} \langle \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top, \mathcal{X}\rangle \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top \\ &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} \mathcal{X}_{ijk} \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top\end{aligned}$$

The proof to such a decomposition is straightforward since $\langle \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top, \mathcal{X}\rangle$ will give out the exact value of $\mathcal{X}_{ijk}$.

The following norms on tensors will be used throughout the proof of our main result.

Definition II.13: (ℓ_{2^*} norm of tensor column): Let $\vec{x}$ be an $n_1 \times 1 \times n_3$ tensor column, we define an ℓ_{2^*} norm on it as follows

$$\|\vec{x}\|_{2^*} = \sqrt{\sum_{i=1}^{n_1}\sum_{k=1}^{n_3} \vec{x}_{i1k}^2}\,. \tag{18}$$

Moreover, we have the following relationship between the ℓ_{2^*} norm of $\vec{x}$ and its FFT along the third dimension $\widehat{\vec{x}}$,

$$\|\vec{x}\|_{2^*} = \frac{1}{\sqrt{n_3}}\|\widehat{\vec{x}}\|_{2^*} \ , \tag{19}$$

where $1/\sqrt{n_3}$ is the normalization constant.

Definition II.14 (Tensor Frobenius norm): The induced Frobenius norm from the inner product defined above is given by,

$$\|\mathcal{A}\|_F = \langle \mathcal{A}, \mathcal{A}\rangle^{1/2} = \frac{1}{\sqrt{n_3}}\|\widehat{\mathcal{A}}\|_F = \sqrt{\sum_i\sum_j\sum_k \mathcal{A}_{ijk}^2}$$

Definition II.15 (Tubal nuclear norm [30]): The tubal nuclear norm of a tensor $\mathcal{A}$, denoted as $\|\mathcal{A}\|_{TNN}$, is the sum of singular values of all the frontal slices of $\widehat{\mathcal{A}}$, and is a convex relaxation of the tensor tubal-rank [8]. In particular,

$$\|\mathcal{A}\|_{TNN} = \|\overline{\mathcal{A}}\|_* \tag{20}$$

Definition II.16 (Tensor spectral norm): The tensor spectral norm $\|\mathcal{A}\|$ of a third-order tensor $\mathcal{A}$ is defined as the largest singular value of $\mathcal{A}$. Moreover,

$$\|\mathcal{A}\| = \|\overline{\mathcal{A}}\| \tag{21}$$

i.e. the tensor spectral norm of $\mathcal{A}$ equals to the matrix spectral norm of $\overline{\mathcal{A}}$.

Definition II.17 (Tensor operator norm): Suppose $\mathcal{L}$ is a tensor operator, then the operator norm of $\mathcal{L}$ is defined as,

$$\|\mathcal{L}\|_{op} = \sup_{\mathcal{X}:\|\mathcal{X}\|_F \le 1} \|\mathcal{L}(\mathcal{X})\|_F \ , \tag{22}$$

which is consistent with matrix case. Spectral norm is equivalent to the operator norm if the tensor operator $\mathcal{L}$ can be represented as a tensor $\mathscr{L}$ t-product $\mathcal{X}$: $\mathcal{L}(\mathcal{X}) = \mathscr{L} * \mathcal{X}$. Then $\|\mathcal{L}\|_{op} = \|\mathscr{L}\|$.

Definition II.18 (Tensor infinity norm): The tensor infinity norm $\|\mathcal{A}\|_\infty$ is defined as follows:

$$\|\mathcal{A}\|_\infty = \max_{i,j,k} |\mathcal{A}_{ijk}| \tag{23}$$

which is the entry with the largest absolute value of $\mathcal{A}$.

III. Main Result

In this section, we will formally define the sampling model and the problem of tensor completion. Our main result is stated in Theorem III.1.

A. Tensor Completion with Random Sampling

Given a third-order tensor $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of tubal-rank r, suppose there are m entries in $\mathcal{M}$ sampled according to the Bernoulli model, which means each entry in the tensor is sampled with probability p independent of the others. The task of tensor completion problem is to recover $\mathcal{M}$ from the observed entries.

In this paper, we follow the approach taken in [8] which solves the following convex optimization problem for tensor completion,

$$\begin{aligned}&\min_{\mathcal{X}} \quad \|\mathcal{X}\|_{TNN} \\ &\text{subject to } \mathcal{X}_{ijk} = \mathcal{M}_{ijk}, (i,j,k) \in \Omega\end{aligned} \tag{24}$$

where Ω is the index set of observed entries. We will analyze the sufficient conditions under which, the solution to (24) is equal to $\mathcal{M}$.

Before we state our main result, we need to introduce the notion of tensor incoherence, a condition that is required for the results to hold true under random sampling. Similar to the matrix completion case, recovery under random sampling is hopeless if most of the entries are equal to zero [9]. For tensor completion using t-SVD, if tensor $\mathcal{M}$ is sparse, then in the reduced t-SVD of $\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, the tensors $\mathcal{U}$ and $\mathcal{V}$ will be highly concentrated on the tensor basis. Similar to the case of matrix completion [9], it is required that the tensor columns, $\mathcal{U}(:, i, :)$ and $\mathcal{V}(:, i, :)$, $i = 1, 2, \ldots, r$ be sufficiently spread, i.e. should be uncorrelated with the tensor basis, for recovery under random sampling. This intuition motivates the following *tensor incoherence condition*.

Definition III.1 (Tensor Incoherence Condition): Let the reduced t-SVD of a tensor $\mathcal{M}$ be $\mathcal{U} * \mathcal{S} * \mathcal{V}^\top$. $\mathcal{M}$ is said to satisfy the *tensor incoherence condition*, if there exists a $\mu_0 > 0$ such that

$$\begin{aligned}\max_{i=1,\ldots,n_1} \left\|\mathcal{U}^\top * \vec{e}_i\right\|_{2^*} &\le \sqrt{\frac{\mu_0 r}{n_1}}, \\ \max_{j=1,\ldots,n_2} \left\|\mathcal{V}^\top * \vec{e}_j\right\|_{2^*} &\le \sqrt{\frac{\mu_0 r}{n_2}},\end{aligned} \tag{25}$$

where $\vec{e}_i$ is the $n_1 \times 1 \times n_3$ column basis with $\vec{e}_{i11} = 1$ and $\vec{e}_j$ is the $n_2 \times 1 \times n_3$ column basis with $\vec{e}_{j11} = 1$

Note that the smallest μ_0 is equal to 1 and this value is achieved when each tensor column $\vec{u}_i = \mathcal{U}(:, i, :)$ has entries with magnitude $1/\sqrt{n_1 n_3}$, or each tensor column $\vec{v}_i = \mathcal{V}(:, i, :)$ has entries with magnitude $1/\sqrt{n_2 n_3}$. The largest possible value of μ_0 is $\min(n_1, n_2)/r$ when one of the tensor columns of $\mathcal{U}$ (or $\mathcal{V}$ respectively) is equal to the tensor column basis $\vec{e}_i$ (or $\vec{e}_j$ resp.). With low μ_0, each entry of $\mathcal{M}$ carries approximately same amount of information.

In [9] for matrix completion case, another joint incoherence condition is needed, which bounds the maximum (absolute value) entry of $UV^\top$, where U and V correspond to left and right singular vectors in the SVD of the matrix. However, this joint incoherence condition is regarded unintuitive and restrictive. [22] successfully removed this joint incoherence by using the $\ell_{\infty,2}$ norm to get a similar bound in the dual certificate step. In our tensor completion case, we apply this idea to our set-up and successfully avoid the joint incoherence condition. Now we will present our main result.

Theorem III.1: Suppose $\mathcal{M}$ is an $n_1 \times n_2 \times n_3$ tensor and its reduced t-SVD is given by $\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$ where $\mathcal{U} \in \mathbb{R}^{n_1 \times r \times n_3}$, $\mathcal{S} \in \mathbb{R}^{r \times r \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times r \times n_3}$. Suppose $\mathcal{M}$ satisfies the tensor incoherence condition(25) with parameter $\mu_0 > 0$, then there exists constants $c_0, c_1, c_2 > 0$ such that if

$$p \geq c_0 \frac{\mu_0 r \log(n_3(n_1+n_2))}{\min\{n_1, n_2\}}, \tag{26}$$

then $\mathcal{M}$ is the unique minimizer to (24) with probability at least $1 - c_1((n_1+n_2)n_3)^{-c_2}$.

Note that the sampling model we use here is the Bernoulli model. There are some other widely used models including sampling with or without replacement. For matrix completion, the recovery guarantees for different models are consistent with only a change of the constants in the sampling complexity and recovery guarantees [31], [32] and we expect them to be the same in our case as well.

Note that although the proof of **Theorem** III.1 follows closely the proof of matrix completion under various measurement models, there are some subtle differences. First of all, we are sampling in the *original* domain, while the tubal nuclear norm (TNN) is defined in the Fourier domain. In fact, let $\mathcal{P}_\Omega(\mathcal{Z})$ denotes the same size tensor as $\mathcal{Z}$ with $\mathcal{P}_\Omega(\mathcal{Z})_{ijk} = \mathcal{Z}_{ijk}$ if $(i,j,k) \in \Omega$ and zero otherwise. Then (24) can be rewritten as,

$$\min_{\mathcal{X}} \quad \|\mathcal{X}\|_{TNN}$$
$$\text{subject to } \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{M}) \ ,$$

Note that it is equivalent to the following

$$\min_{\mathcal{X}} \quad \|\overline{\mathcal{X}}\|_*$$
$$\text{subject to } \mathcal{F}\mathcal{P}\mathcal{F}^{-1}(\overline{\mathcal{X}}) = \mathcal{F}\mathcal{P}\mathcal{F}^{-1}(\overline{\mathcal{M}}) \ , \tag{27}$$

where $\mathcal{F}$ is a mapping which maps a third order tensor $\mathcal{Z}$ to $\overline{\mathcal{Z}}$, and $\mathcal{F}^{-1}$ is its inverse transform. Now the above problem is a matrix completion problem under linear constraint and it is completely defined in the Fourier domain.

We can regard this constraint as several observations in the form of inner products of $\overline{\mathcal{X}}$ and $\overline{\mathcal{E}_{ijk}}$ in the Fourier domain, where $\mathcal{E}_{ijk}$ is a unit tensor with only (i,j,k)th entry being 1 and $\overline{\mathcal{E}_{ijk}} = \mathtt{fft}(\mathcal{E}_{ijk}, [\,], 3)$. In details, the problem can be equivalently rewritten as,

$$\min \quad \|\overline{\mathcal{X}}\|_*$$
$$\text{subject to} \quad \langle \overline{\mathcal{X}}, \overline{\mathcal{E}_{ijk}} \rangle = \langle \overline{\mathcal{M}}, \overline{\mathcal{E}_{ijk}} \rangle, \quad (i,j,k) \in \Omega \tag{28}$$
$$\overline{\mathcal{X}} \text{ is block diagonal (block size } n \times n).$$

We now note that [33] gives provable guarantees on recovery for such cases, but without the block diagonal constraint, provided that, (a) the basis are orthonormal, and (B) the basis and matrix satisfy a standard incoherence condition and an extra joint incoherence condition.

However, because of the block diagonal constraint, the result of [33] cannot be directly applied to derive provable bounds to our problem even if we are given an extra joint incoherence condition. In fact, following this formula we can also change the block diagonal constraint in (28) into some more inner product observations, which force the entries outside the block diagonal positions to be zero. But this actually needs to observe even more entries ($O(n^2 n_3^2)$) than the data tensor itself to finish the recovery. On the other hand, we also cannot solve (28) slice-wise (which avoids the block diagonal constraint) using matrix completion results, since the constraint $\langle \overline{\mathcal{X}}, \overline{\mathcal{E}_{ijk}} \rangle = \langle \overline{\mathcal{M}}, \overline{\mathcal{E}_{ijk}} \rangle$ is not separable on each frontal slice. As we will see in the next section, in the case of *random tubal sampling*, the tensor completion problem can be separated into individual matrix completion problem on each frontal slice, with random sampling in the Fourier domain.

B. Tensor Completion with Random Tubal Sampling

Another way to sample a tensor is to perform random or adaptive tubal sampling as considered in [34] for fingerprinting application. Here we will comment on the theoretical guarantees for recovery under random tubal sampling, since [34] did not specifically analyzed this. Instead of randomly sampling entries of a third-order tensor $\mathcal{M}$ as in the previous subsection, one can randomly sample tensor tubes along the third dimension. Then the completion problem becomes,

$$\min_{\mathcal{X}} \quad \|\mathcal{X}\|_{TNN}$$
$$\text{subject to } \mathcal{X}_{ijk} = \mathcal{M}_{ijk}, (i,j) \in \Omega, k = 1, 2, \ldots, n_3, \tag{29}$$

where Ω is the index set of observed tubes. If we take FFT of $\mathcal{X}$ and $\mathcal{M}$ along the third dimension, it is easy to see that solving the above optimization problem is equivalent to solving n_3 matrix completion problems in the Fourier domain,

$$\min_{\widehat{\mathcal{X}}^{(k)}} \quad \|\widehat{\mathcal{X}}^{(k)}\|_*$$
$$\text{subject to } \widehat{\mathcal{X}}_{ij}^{(k)} = \widehat{\mathcal{M}}_{ij}^{(k)}, (i,j) \in \Omega \tag{30}$$

for $k = 1, 2, \ldots, n_3$. Therefore tensor completion problem with tubal sampling is essentially the matrix completion from random samplings of each frontal slice in the Fourier domain. Then we can directly use the result of matrix completion here.

Suppose there are p third-dimensional tubes of $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ sampled according to the Bernoulli model, which means each tube in the tensor is sampled with probability p independent of the other tubes. Then we have the following theorem,

Theorem III.2: [34] Let $\mathcal{M}$ be an $n_1 \times n_2 \times n_3$ tensor and its reduced t-SVD is given by $\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$ where $\mathcal{U} \in \mathbb{R}^{n_1 \times r \times n_3}$, $\mathcal{S} \in \mathbb{R}^{r \times r \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times r \times n_3}$. Suppose each frontal slice $\widehat{\mathcal{M}}^{(i)}$ satisfies the matrix weak incoherence condition (31) with parameter $\mu_0 > 0$. Then there exists constants $c_0, c_1, c_2 > 0$ such that if

$$p \geq c_0 \frac{\mu_0 r \log^2(n_1+n_2)}{\min\{n_1, n_2\}},$$

then $\mathcal{M}$ is the unique minimizer to (24) with probability at least $1 - c_1 n_3 (n_1+n_2)^{-c_2}$.

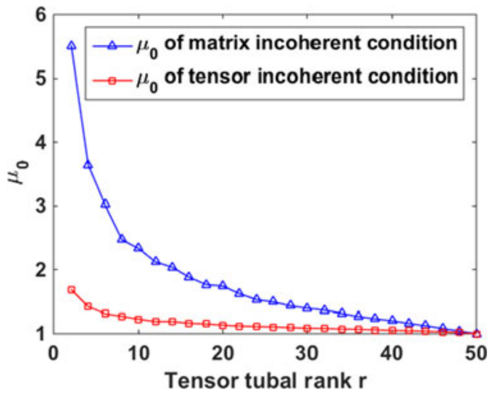


Fig. 5. Comparison of tensor and matrix incoherent condition on $50 \times 50 \times 20$ tensor.

Remark III.1 (A comment on the incoherence conditions for tubal sampling): In Theorem III.2 each slice $\widehat{\mathcal{M}}^{(k)}$ in the Fourier domain needs to satisfy *matrix weak incoherence condition* [22] with parameter μ_0. That is for all $k \in \{1, 2, \dots, n_3\}$,

$$\max_{i=1,2,..,n_1} \|\widehat{\mathcal{U}}^{(k)\top} \mathbf{e}_i\|_2 \leq \sqrt{\frac{\mu_0 r}{n_1}},$$
$$\max_{j=1,2,..,n_2} \|\widehat{\mathcal{V}}^{(k)\top} \mathbf{e}_j\|_2 \leq \sqrt{\frac{\mu_0 r}{n_2}}, \tag{31}$$

where $\mathbf{e}_i$ denotes the i-th standard basis in $\mathbb{R}^n$. Note that we have $1 \leq \mu_0 \leq \min\{n_1, n_2\}/r$. We now show that the matrix incoherence condition of Equation (31) is not equivalent to the tensor incoherence conditions of Equations (25). In fact from Equations (31) we have,

$$\max_{i\in\{1,2,..,n_1\}} \|\widehat{\mathcal{U}}^{(k)\top} \mathbf{e}_i\|_2 \leq \sqrt{\frac{\mu_0 r}{n_1}} \tag{32}$$
$$\Longrightarrow \max_{i=1,2,..,n_1} \sum_{k=1}^{n_3} \|\widehat{\mathcal{U}}^{(k)\top} \mathbf{e}_i\|_2^2 \leq \frac{n_3 \mu_0 r}{n_1}$$
$$\Longleftrightarrow \max_{i=1,2,..,n_1} \|\widehat{\mathcal{U}}_i^{\top}\|_{2^*}^2 \leq \frac{n_3 \mu_0 r}{n_1}$$
$$\Longleftrightarrow \max_{i=1,2,..,n_1} \left\|\mathcal{U}^\top * \vec{e}_i\right\|_{2^*} \leq \sqrt{\frac{\mu_0 r}{n_1}}. \tag{33}$$

Similarly we can get,

$$\max_{j=1,\dots,n_2} \left\|\mathcal{V}^\top * \vec{e}_j\right\|_{2^*} \leq \sqrt{\frac{\mu_0 r}{n_2}} ,$$

which is exactly the tensor incoherence condition. Therefore our tensor incoherence condition can be obtained from the matrix incoherence condition, but not vice versa since (33) does not imply (32).

Figure 5 shows a comparison of tensor and matrix incoherence condition. Each time we randomly generate a $50 \times 50 \times 20$ tensor with different tubal-rank and compute μ_0 of both cases. We repeat this process 20 times and plot the average μ_0 in the figure. As we can tell, μ_0 of tensor incoherence condition is indeed lower than that of matrix incoherence condition. In the worst case, when tensor tubal-rank is small, the two incoherence constants can differ by an order of $\sqrt{n_3}$.

IV. Proof of Theorem III.1

In this section, we provide the proof of Theorem III.1. The main idea is to use convex analysis to derive conditions under which one can verify whether $\mathcal{M}$ is the unique minimum tubal nuclear norm solution to (24), and then to explicitly show that such conditions are met with high probability under the conditions of Theorem III.1.

To simplify the notation and without loss of generality we assume $n_1 = n_2 = n$ and do not put any assumption on n_3.

Before continuing, some notations used in the proof should be clarified. A tensor $\mathcal{Y}$ is the subgradient of $\|\cdot\|_{TNN}$ at $\mathcal{M}_0$ (denoted $\mathcal{Y} \in \partial\|\mathcal{M}_0\|_{TNN}$), if for all $\mathcal{X} \in \mathbb{R}^{n\times n\times n_3}$,

$$\|\overline{\mathcal{X}}\|_* \geq \|\overline{\mathcal{M}}_0\|_* + \langle \overline{\mathcal{Y}}, \overline{\mathcal{X}} - \overline{\mathcal{M}}_0\rangle . \tag{34}$$

Recall that a matrix $\mathbf{Y}$ is a subgradient of a convex function $f : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}$ at matrix $\mathbf{M}_0$ if

$$f(\mathbf{X}) \geq f(\mathbf{M}_0) + \langle \mathbf{Y}, \mathbf{X} - \mathbf{M}_0\rangle .$$

Moreover, $\mathbf{Y}$ is a subgradient of the nuclear norm at $\mathbf{M}_0$ if and only if $\mathbf{Y}$ is of the form [35]

$$\mathbf{Y} = \mathbf{U}\mathbf{V}^\top + \mathbf{W} , \tag{35}$$

where $\mathbf{U}\mathbf{S}_0\mathbf{V}^\top = \text{svd}(\mathbf{M}_0)$ is the singular value decomposition of $\mathbf{M}_0$, and $\mathbf{W}$ satisfies

1) $\mathbf{U}^\top\mathbf{W} = 0, \quad \mathbf{W}\mathbf{V} = 0$
2) $\|\mathbf{W}\| \leq 1$.

Similarly, let $\mathcal{Y} = \mathcal{U} * \mathcal{V}^\top + \mathcal{W}$, where $\mathcal{U} * \mathcal{S}_0 * \mathcal{V}^\top$ is the t-SVD of $\mathcal{M}_0$, and $\|\mathcal{W}\| \leq 1$. One can verify that such a $\mathcal{Y}$ satisfies (34), therefore $\mathcal{Y} \in \partial\|\mathcal{M}_0\|_{TNN}$.

In order to proceed, we introduce the orthogonal decomposition $\mathbb{R}^{n\times n\times n_3} = T \oplus T^\perp$, where T is the linear space spanned by the elements of the form $\mathcal{U}(:, k, :) * \vec{x}^\top$ and $\vec{y} * \mathcal{V}(:, k, :)^\top$, where $\vec{x}, \vec{y} \in \mathbb{R}^{n\times 1\times n_3}$ are arbitrary tensor columns, $k = 1, 2, \dots, r$. Let $T^\perp$ be its orthogonal complement. The orthogonal projections $\mathcal{P}_T$ onto T and $\mathcal{P}_{T^\perp}$ onto $T^\perp$ are given as follows,

$$\mathcal{P}_T(\mathcal{Z}) = \mathcal{U} * \mathcal{U}^\top * \mathcal{Z} + \mathcal{Z} * \mathcal{V} * \mathcal{V}^\top - \mathcal{U} * \mathcal{U}^\top * \mathcal{Z} * \mathcal{V} * \mathcal{V}^\top \tag{36}$$

$$\mathcal{P}_{T^\perp}(\mathcal{Z}) = \mathcal{Z} - \mathcal{P}_T(\mathcal{Z}) = (\mathcal{I} - \mathcal{U} * \mathcal{U}^\top) * \mathcal{Z} * (\mathcal{I} - \mathcal{V} * \mathcal{V}^\top) \tag{37}$$

where $\mathcal{I}$ is the identity tensor of size $n \times n \times n_3$.

Define a random variable $\delta_{ijk} = \mathbf{1}_{(i,j,k)\in\Omega}$ where $\mathbf{1}_{(\cdot)}$ is the indicator function. Let $\mathcal{R}_\Omega : \mathbb{R}^{n\times n\times n_3} \to \mathbb{R}^{n\times n\times n_3}$ be a random projection as follows,

$$\mathcal{R}_\Omega(\mathcal{Z}) = \sum_{i,j,k} \frac{1}{p} \delta_{ijk} \mathcal{Z}_{ijk} \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top . \tag{38}$$

Similar to the matrix completion case, in the following we will construct a *dual certificate* $\mathcal{Y}$ and show that it is close to the subgradient of $\|\mathcal{M}\|_{TNN}$ under certain conditions. The following Proposition IV.1 and Lemma IV.1 directly support the proof of the main theorem. The proofs of these technical and supporting results are provided in the Appendices A-D.

Proposition IV.1: Tensor $\mathcal{M}$ is the unique minimizer to (24) if the following conditions hold:

1) $\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\|_{op} \le \frac{1}{2}$
2) There exists a tensor $\mathcal{Y}$ such that $\mathcal{P}_\Omega(\mathcal{Y}) = \mathcal{Y}$ and
   a) $\|\mathcal{P}_T(\mathcal{Y}) - \mathcal{U} * \mathcal{V}^\top\|_F \le \frac{1}{4nn_3^2}$
   b) $\|\mathcal{P}_{T^\perp}(\mathcal{Y})\| \le \frac{1}{2}$

Lemma IV.1: Suppose $\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\|_{op} \le \frac{1}{2}$. Then for any $\mathcal{Z}$ such that $\mathcal{P}_\Omega(\mathcal{Z}) = 0$, we have

$$\frac{1}{2}\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{Z})\|_{TNN} > \frac{1}{4nn_3}\|\mathcal{P}_T(\mathcal{Z})\|_F\,. \tag{39}$$

Proof of Proposition IV.1: The main idea is that we want to prove that for any $\mathcal{Z}$ supported in Ω^c, $\|\mathcal{M} + \mathcal{Z}\|_{TNN} > \|\mathcal{M}\|_{TNN}$. To prove this the following three facts are used.

Fact IV.1: $\|\mathcal{A}\|_{TNN} = n_3 \sup_{\|\mathcal{B}\|\le 1}\langle \mathcal{A}, \mathcal{B}\rangle$, where $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n\times n\times n_3}$. Specifically, if the t-SVD of $\mathcal{A}$ is given by $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, then let $\mathcal{B} = \mathcal{U} * \mathcal{V}^\top$. Obviously $\|\mathcal{B}\| \le 1$ and we have $n_3\langle\mathcal{A}, \mathcal{B}\rangle = \text{trace}(\overline{\mathcal{S}}) = \|\mathcal{A}\|_{TNN}$.

Recall that for matrix case, we have $\|\mathbf{A}\|_* = \sup_{\|\mathbf{B}\|\le 1}\langle \mathbf{A}, \mathbf{B}\rangle$, where $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$. Then the fact comes from,

$$\|\mathcal{A}\|_{TNN} = \|\overline{\mathcal{A}}\|_* = \sup_{\|\overline{\mathcal{B}}\|\le 1}\langle\overline{\mathcal{A}}, \overline{\mathcal{B}}\rangle$$
$$= n_3 \sup_{\|\mathcal{B}\|\le 1}\langle\mathcal{A}, \mathcal{B}\rangle\,.$$

Define the t-SVD of $\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{Z})$ to be $\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{Z}) = \mathcal{U}_\perp * \mathcal{S}_\perp * \mathcal{V}_\perp^\top$, where $\mathcal{Z} \in \mathbb{R}^{n\times n\times n_3}$ such that $\mathcal{P}_\Omega(\mathcal{Z}) = 0$. Then use the fact above we have

$$\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{Z})\|_{TNN} = n_3\langle\mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{P}_{\mathcal{T}^\perp}(\mathcal{Z})\rangle\,. \tag{40}$$

Fact IV.2: $\|\mathcal{M}\|_{TNN} = n_3\langle\mathcal{U} * \mathcal{V}^\top + \mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{M}\rangle$

Since $\mathcal{P}_T(\mathcal{U}) = \mathcal{U}$ and $\mathcal{P}_{T^\perp}(\mathcal{U}_\perp) = \mathcal{U}_\perp$, we have $\mathcal{U} * \mathcal{U}_\perp^\top = 0$ and similarly $\mathcal{V} * \mathcal{V}_\perp^\top = 0$ by definition. Then the fact can be verified by the following.

$$\begin{aligned}
&n_3\langle\mathcal{U} * \mathcal{V}^\top + \mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{M}\rangle \\
&= n_3\langle\mathcal{U} * \mathcal{V}^\top + \mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{U} * \mathcal{S} * \mathcal{V}^\top\rangle \\
&= \text{trace}((\overline{\mathcal{U}}\overline{\mathcal{V}}^\top + \overline{\mathcal{U}}_\perp\overline{\mathcal{V}}_\perp^\top)^\top\, \overline{\mathcal{U}}\overline{\mathcal{S}}\overline{\mathcal{V}}^\top) \\
&= \text{trace}(\overline{\mathcal{S}}) = \|\mathcal{M}\|_{TNN}
\end{aligned}$$

Fact IV.3: $\|\mathcal{U} * \mathcal{V}^\top + \mathcal{U}_\perp * \mathcal{V}_\perp^\top\| = 1$

Consider a matrix Q such that

$$Q = \overline{\mathcal{U}}\overline{\mathcal{V}}^\top + \overline{\mathcal{U}}_\perp\overline{\mathcal{V}}_\perp^\top = [\overline{\mathcal{U}} \quad \overline{\mathcal{U}}_\perp]\begin{bmatrix}\overline{\mathcal{V}}^\top \\ \overline{\mathcal{V}}_\perp^\top\end{bmatrix}$$

Since $\overline{\mathcal{U}}^\top\overline{\mathcal{U}}_\perp = 0$ and $\overline{\mathcal{V}}^\top\overline{\mathcal{V}}_\perp = 0$, the above expression is the matrix singular value decomposition of Q, so we have

$$\|\mathcal{U} * \mathcal{V}^\top + \mathcal{U}_\perp * \mathcal{V}_\perp^\top\| = \|\overline{\mathcal{U}}\overline{\mathcal{V}}^\top + \overline{\mathcal{U}}_\perp\overline{\mathcal{V}}_\perp^\top\|$$
$$= \|Q\| = 1$$

Now using the above facts, given any $\mathcal{Z} \in \mathbb{R}^{n\times n\times n_3}$ such that $\mathcal{P}_\Omega(\mathcal{Z}) = 0$, we have

$$\begin{aligned}
&\|\mathcal{M} + \mathcal{Z}\|_{TNN} \\
&\ge n_3\langle\mathcal{U} * \mathcal{V}^\top + \mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{M} + \mathcal{Z}\rangle && (41)\\
&= \|\mathcal{M}\|_{TNN} + n_3\langle\mathcal{U} * \mathcal{V}^\top + \mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{Z}\rangle \\
&= \|\mathcal{M}\|_{TNN} + n_3\langle\mathcal{U} * \mathcal{V}^\top, \mathcal{P}_T(\mathcal{Z})\rangle \\
&\quad + n_3\langle\mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{P}_{T^\perp}(\mathcal{Z})\rangle \\
&= \|\mathcal{M}\|_{TNN} + n_3\langle\mathcal{U} * \mathcal{V}^\top, \mathcal{P}_T(\mathcal{Z})\rangle \\
&\quad + n_3\langle\mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{P}_{T^\perp}(\mathcal{Z})\rangle - n_3\langle\mathcal{Y}, \mathcal{Z}\rangle && (42)\\
&= \|\mathcal{M}\|_{TNN} + n_3\langle\mathcal{U} * \mathcal{V}^\top - \mathcal{P}_T(\mathcal{Y}), \mathcal{P}_T(\mathcal{Z})\rangle \\
&\quad + n_3\langle\mathcal{U}_\perp * \mathcal{V}_\perp^\top - \mathcal{P}_{T^\perp}(\mathcal{Y}), \mathcal{P}_{T^\perp}(\mathcal{Z})\rangle \\
&= \|\mathcal{M}\|_{TNN} + \langle\overline{\mathcal{U}}\overline{\mathcal{V}}^\top - \overline{\mathcal{P}_T(\mathcal{Y})}, \overline{\mathcal{P}_T(\mathcal{Z})}\rangle \\
&\quad + \|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{TNN} - \langle\overline{\mathcal{P}_{T^\perp}(\mathcal{Y})}, \overline{\mathcal{P}_{T^\perp}(\mathcal{Z})}\rangle && (43)\\
&\ge \|\mathcal{M}\|_{TNN} - \|\overline{\mathcal{U}}\overline{\mathcal{V}}^\top - \overline{\mathcal{P}_T(\mathcal{Y})}\|_F\|\overline{\mathcal{P}_T(\mathcal{Z})}\|_F \\
&\quad + \|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{TNN} - \|\overline{\mathcal{P}_{T^\perp}(\mathcal{Y})}\|\|\overline{\mathcal{P}_{T^\perp}(\mathcal{Z})}\|_* && (44)\\
&= \|\mathcal{M}\|_{TNN} - n_3\|\mathcal{U} * \mathcal{V}^\top - \mathcal{P}_T(\mathcal{Y})\|_F\|\mathcal{P}_T(\mathcal{Z})\|_F \\
&\quad + \|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{TNN} - \|\mathcal{P}_{T^\perp}(\mathcal{Y})\|\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{TNN} \\
&\ge \|\mathcal{M}\|_{TNN} - \frac{1}{4nn_3}\|\mathcal{P}_T(\mathcal{Z})\|_F + \frac{1}{2}\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{TNN} \\
&> \|\mathcal{M}\|_{TNN} && (45)
\end{aligned}$$

where (41) uses the Fact IV.1; $\mathcal{Y}$ in (42) is a tensor dual certificate supported in Ω such that $\mathcal{P}_\Omega(\mathcal{Y}) = \mathcal{Y}$. So it is easy to show $\langle\mathcal{Z}, \mathcal{Y}\rangle = 0$ using the standard basis decomposition; (43) uses equation (40); (44) is based on the following two facts for any same size matrices $\mathbf{A}$ and $\mathbf{B}$,

$$|\langle\mathbf{A}, \mathbf{B}\rangle| \le \|\mathbf{A}\|_F\|\mathbf{B}\|_F$$
$$\langle\mathbf{A}, \mathbf{B}\rangle \le \|\mathbf{A}\|\|\mathbf{B}\|_*\,,$$

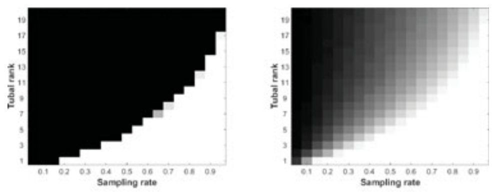
and (45) uses the Condition **2** of Proposition IV.1.

Therefore, for any $\mathcal{X} \ne \mathcal{M}$ obeying $\mathcal{P}_\Omega(\mathcal{X} - \mathcal{M}) = 0$, we have $\|\mathcal{X}\|_{TNN} > \|\mathcal{M}\|_{TNN}$, which proves $\mathcal{M}$ is the unique minimizer of (24).

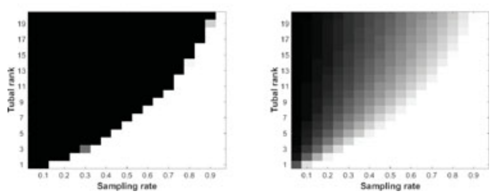
Proof of Theorem III.1: As proved in the Appendix A, B and C, if p satisfies (26), the Condition 1 and Condition 2 in Proposition IV.1 are satisfied with probability at least $1 - c_1((n_1 + n_2)n_3)^{-c_2}$ for some positive constants c_1 and c_2. The proof of Theorem III.1 then follows directly from Proposition IV.1, which states that $\mathcal{M}$ is the unique minimizer to (24).

V. Numerical Experiments

To demonstrate our results, we conducted some numerical experiments to recover third order tensors of different sizes and tubal-ranks r from m observed entries. For each episode we generated an $n_1 \times n_2 \times n_3$ random tensor with i.i.d. Gaussian entries, performed the t-SVD of it, kept the first r singular tubes

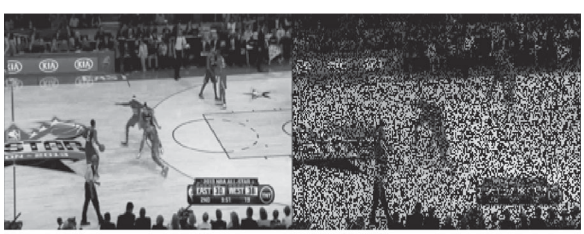


(a) $40 \times 40 \times 30$ tensor.

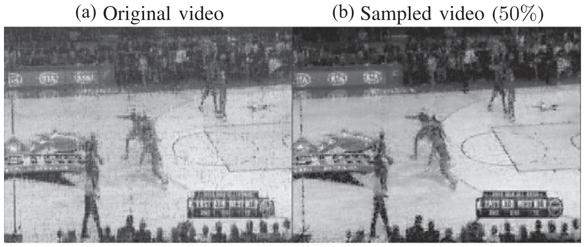


(b) $30 \times 30 \times 20$ tensor.

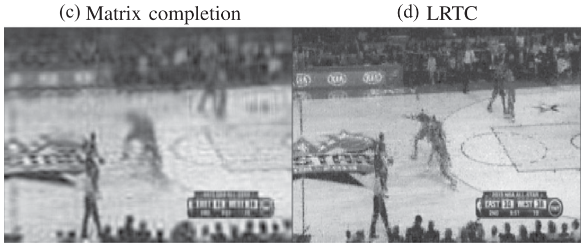
Fig. 6. **Recovery of third order tensors from their entries.** In the left figures of both cases, each cell's value reflects the empirical recovery rate. Black denotes failure and white denotes success in recovery in all simulations. In the right figures of both cases, each cell's value is the RSE of the recovery under the corresponding sampling rate and tubal rank. Black denotes 1 and white denotes 0.



(a) Original video (b) Sampled video (50%)



(c) Matrix completion (d) LRTC



(e) geomCG (f) Our method

Fig. 7. The 50th frame of tensor completion result on a basketball video.

TABLE II
RSE OF TENSOR COMPLETION RESULT ON THE BASKETBALL VIDEO

Completion Approach	RSE(dB)
Matrix completion on each frame	-17.02
LRTC	-18.29
Tensor completion in Tucker format by Riemannian optimization	-16.47
Tensor completion with random sampling via tSVD	-20.77

and got $\mathcal{M}$ with tubal rank r. We sampled m entries of $\mathcal{M}$ uniformly at random and try to recover $\mathcal{M}$ using (24). We denote the solution of (24) by $\mathcal{X}$ and compute the relative square error (RSE): $\|\mathcal{X} - \mathcal{M}\|_F / \|\mathcal{M}\|_F$. If the RSE $\leq 10^{-3}$, then we claim that the recovery is exact. We repeated our experiments 20 times. The results are shown in Figure 6. In the left figures, the color of each cell reflects the empirical recovery rate ranging from 0 to 1. White cell means exact recoveries in all experiments; and black cell means all experiments failed. The right figures are the RSE plots of one typical run of the simulation. For each cell, the value reflects the RSE of the recovery under the corresponding sampling rate and tubal rank. Black denotes 1 and white is 0.

For practical application, we tested 4 algorithms for video data completion from randomly missing entries. The first method is performing matrix completion [36] on each frame of the tensor; the second approach is the Low Rank Tensor Completion (LRTC) algorithm in [7]; the third one is the tensor completion in Tucker format by Riemannian optimization (geomCG) [37]; and the last approach is our tensor completion with random sampling via t-SVD. The size of the video is $144 \times 256 \times 80$ and we randomly sampled 50% entries from the video.

The result is shown in Figure 7. We compared relative square error (RSE) of each approach in our simulation and the result is in Table II from which we can see that our approach yields a better performance over the other 3 methods.

For more numerical results for completion using the t-SVD, we refer the reader to our paper [8].

VI. CONCLUSION

In this paper we considered the problem of recovering third-order tensors under random sampling. Using a recently proposed algebraic framework and a notion of rank, namely the tubal-rank, we show that under the certain tensor incoherence conditions, one can exactly recover a third-order tensor with low tubal-rank, and establish a theoretical bound for exact completion when using a convex optimization algorithm for recovery. We compare different tensor completion approaches experimentally and show that our method yields a better performance over the others on some real data sets.

Our results indicate that the algebraic framework of Section II can provide a computationally efficient as well as theoretically sound framework to model multidimensional data. Moreover, in this framework the tools and methods developed to deal with prediction and learning problems using the vector space methods can be readily adapted. We will explore this connection more concretely in our future work.

APPENDIX A
PROOF OF PROPOSITION IV.1 CONDITION 1

The following theorem is first developed in [10], and it will be used frequently in this section.

Theorem VII.1 (Noncommutative Bernstein Inequality): Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_L$ be independent zero-mean random matrices of dimension $d_1 \times d_2$. Suppose

$$\rho_k^2 = \max\{\|\mathbb{E}[\mathbf{X}_k \mathbf{X}_k^\top]\|, \|\mathbb{E}[\mathbf{X}_k^\top \mathbf{X}_k]\|\}$$

and

$$\|\mathbf{X}_k\| \leq M$$

almost surely for all k. Then for any $\tau > 0$,

$$\mathbb{P}\left[\left\|\sum_{k=1}^{L}\mathbf{X}_k\right\| > \tau\right] \leq (d_1+d_2)\exp\left(\frac{-\tau^2/2}{\sum_{k=1}^{L}\rho_k^2 + M\tau/3}\right) \tag{46}$$

This theorem is a corollary of a Chernoff bound for finite dimension operators developed by [38]. An extension of this theorem [39] states that if

$$\max\left\{\left\|\sum_{k=1}^{L}\mathbf{X}_k\mathbf{X}_k^\top\right\|, \left\|\sum_{k=1}^{L}\mathbf{X}_k^\top\mathbf{X}_k\right\|\right\} \leq \sigma^2 \tag{47}$$

and let

$$\tau = \sqrt{4c\sigma^2\log(d_1+d_2)} + cM\log(d_1+d_2)$$

for any $c > 0$. Then (46) becomes

$$\mathbb{P}\left[\left\|\sum_{k=1}^{L}\mathbf{X}_k\right\| \geq \tau\right] \leq (d_1+d_2)^{-(c-1)} \tag{48}$$

The following fact is used frequently in this section.

Fact A.1: $\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\|_F^2 \leq \frac{2\mu_0 r}{n}$

Proof of Fact A.1: Following the definition of $\mathcal{P}_T$ in (36), we have

$$\begin{aligned}&\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\\ &= \mathcal{U}*\mathcal{U}^\top * \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top + \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top * \mathcal{V}*\mathcal{V}^\top\\ &\quad - \mathcal{U}*\mathcal{U}^\top * \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top * \mathcal{V}*\mathcal{V}^\top\end{aligned}$$

This gives

$$\begin{aligned}&\left\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\right\|_F^2 = \langle \mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top), \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top\rangle\\ &= \|\mathcal{U}*\mathcal{U}^\top * \vec{e}_i\|_F^2 + \|\mathcal{V}*\mathcal{V}^\top * \vec{e}_j\|_F^2\\ &- \|\mathcal{U}*\mathcal{U}^\top * \vec{e}_i\|_F^2\|\mathcal{V}*\mathcal{V}^\top * \vec{e}_j\|_F^2\\ &\leq \mu_0 r\frac{n_1+n_2}{n_1n_2} \leq \frac{2\mu_0 r}{n}\end{aligned}$$

where the first inequality comes from the tensor incoherence condition (25).

Proof of Proposition IV.1 Condition (1):

Proof: First note that

$$\mathbb{E}[\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T] = \mathcal{P}_T(\mathbb{E}\mathcal{R}_\Omega)\mathcal{P}_T = \mathcal{P}_T\ , \tag{49}$$

which gives

$$\mathbb{E}[\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T] = 0$$

and

$$\mathbb{E}[\overline{\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T} - \overline{\mathcal{P}_T}] = 0 \tag{50}$$

Our goal is to prove the operator $\overline{\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T}$ is not far away from its expected value $\overline{\mathcal{P}_T}$ in the spectral norm using the Non-commutative Bernstein Inequality.

Given any tensor $\mathcal{Z}$ of size $n\times n\times n_3$, we can decompose $\mathcal{P}_T(\mathcal{Z})$ as the following

$$\begin{aligned}\mathcal{P}_T(\mathcal{Z}) &= \sum_{i,j,k}\langle \mathcal{P}_T(\mathcal{Z}), \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top\rangle \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top\\ &= \sum_{i,j,k}\langle \mathcal{Z}, \mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\rangle \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top\end{aligned}$$

This gives

$$\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z}) = \sum_{i,j,k}\frac{1}{p}\delta_{ijk}\langle \mathcal{Z}, \mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\rangle \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top$$

and

$$\begin{aligned}\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z}) = &\sum_{i,j,k}\frac{1}{p}\delta_{ijk}\langle \mathcal{Z}, \mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\rangle\\ &\times \mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\end{aligned}$$

which implies

$$\begin{aligned}\overline{\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T}(\overline{\mathcal{Z}}) = &\sum_{i,j,k}\frac{1}{p}\delta_{ijk}\langle \mathcal{Z}, \mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\rangle\\ &\times \overline{\mathcal{P}_T}(\overline{\vec{e}_i \mathring{e}_k \vec{e}_j^\top})\end{aligned}$$

Define operator $\mathcal{T}_{ijk}$ which maps $\mathcal{Z}$ to $\frac{1}{p}\delta_{ijk}\langle \mathcal{Z}, \mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\rangle \mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)$. Observe that $\|\mathcal{T}_{ijk}\|_{op} = \|\overline{\mathcal{T}_{ijk}}\| = \frac{1}{p}\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\|_F^2$ and $\|\mathcal{P}_T\|_{op} = \|\overline{\mathcal{P}_T}\| \leq 1$. Then we have

$$\begin{aligned}&\|\mathcal{T}_{ijk} - \frac{1}{n^2n_3}\mathcal{P}_T\|_{op} = \|\overline{\mathcal{T}_{ijk}} - \frac{1}{n^2n_3}\overline{\mathcal{P}_T}\|\\ &\leq \max\left\{\frac{1}{p}\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\|_F^2, \frac{1}{n^2n_3}\right\} \leq \frac{2\mu_0 r}{np}\end{aligned}$$

where the first inequality uses the fact that if A and B are positive semidefinite matrices, then $\|A - B\| \leq \max\{\|A\|, \|B\|\}$.

On the other hand, observe that from (49) we have $\mathbb{E}[\mathcal{T}_{ijk}] = \frac{1}{n^2n_3}\mathcal{P}_T$. So

$$\begin{aligned}&\left\|\mathbb{E}[(\overline{\mathcal{T}_{ijk}} - \frac{1}{n^2n_3}\overline{\mathcal{P}}_T)^2]\right\| = \left\|\mathbb{E}[(\mathcal{T}_{ijk} - \frac{1}{n^2n_3}\mathcal{P}_T)^2]\right\|_{op}\\ &= \left\|\mathbb{E}[\frac{1}{p}\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\|_F^2\mathcal{T}_{ijk}] - \frac{2}{n^2n_3}\mathcal{P}_T\mathbb{E}[\mathcal{T}_{ijk}]\right.\\ &\quad\left. + \frac{1}{n^4n_3^2}\mathcal{P}_T\right\|_{op}\\ &= \left\|\frac{1}{p}\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\|_F^2\frac{1}{n^2n_3}\mathcal{P}_T - \frac{1}{n^4n_3^2}\mathcal{P}_T\right\|_{op}\\ &< \left(\frac{1}{p}\frac{2\mu_0 r}{n}\frac{1}{n^2n_3}\right)\|\mathcal{P}_T\|_{op} \leq \frac{2\mu_0 r}{n^3n_3p}\end{aligned}$$

Now let

$$\tau = \sqrt{\frac{14\beta\mu_0 r\log(nn_3)}{3np}} \leq \frac{1}{2}$$

with some constant $\beta > 1$. The inequality holds given p satisfying (26) with c_0 large enough. Use **Theorem** A.1 we have

$$\begin{aligned}
&\mathbb{P}\left[\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\|_{op} > \tau\right] \\
&= \mathbb{P}\left[\|\overline{\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T} - \overline{\mathcal{P}_T}\| > \tau\right] \\
&= \mathbb{P}\left[\left\|\sum_{i,j,k}\left(\overline{\mathcal{T}_{ijk}} - \frac{1}{n^2n_3}\overline{\mathcal{P}_T}\right)\right\| > \tau\right] \\
&\leq 2nn_3 \exp\left(\frac{-\frac{7}{3}\frac{\beta\mu_0 r\log(nn_3)}{np}}{\frac{2\mu_0 r}{np} + \frac{2\mu_0 r}{np}\frac{1}{6}}\right) \leq 2(nn_3)^{1-\beta}
\end{aligned}$$

Therefore we can get

$$\begin{aligned}
&\mathbb{P}\left[\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\|_{op} \leq \frac{1}{2}\right] \\
&\geq \mathbb{P}\left[\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\|_{op} \leq \tau\right] \geq 1 - 2(nn_3)^{1-\beta},
\end{aligned}$$

which finishes the proof. ∎

Appendix B
Proof of Lemma IV.1

Proof: Given any $\mathcal{Z}$ such that $\mathcal{P}_\Omega(\mathcal{Z}) = 0$ and $\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\|_{op} \geq 1/2$, we have

$$\langle\overline{\mathcal{Z}}, \overline{\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z})} - \overline{\mathcal{P}_T(\mathcal{Z})}\rangle \geq -\frac{1}{2}\|\overline{\mathcal{Z}}\|_F$$

which gives

$$\langle\mathcal{Z}, \mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z}) - \mathcal{P}_T(\mathcal{Z})\rangle \geq -\frac{1}{2}\|\mathcal{Z}\|_F$$

Note that

$$\begin{aligned}
&\langle\mathcal{Z}, \mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z})\rangle = \frac{1}{\sqrt{n_3}}\langle\overline{\mathcal{Z}}, \overline{\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z})}\rangle \\
&= \frac{1}{\sqrt{n_3}}\|\overline{\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z})}\|_F^2 = \sqrt{n_3}\|\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z})\|_F^2 \\
&= \sqrt{n_3}\|\mathcal{R}_\Omega(\mathcal{Z} - \mathcal{P}_{T^\perp}(\mathcal{Z}))\|_F^2 = \sqrt{n_3}\|\mathcal{R}_\Omega\mathcal{P}_{T^\perp}(\mathcal{Z})\|_F^2 \\
&\leq \frac{\sqrt{n_3}}{p^2}\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_F^2
\end{aligned}$$

Thus

$$\begin{aligned}
&\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_F^2 \geq \frac{p^2}{\sqrt{n_3}}\langle\mathcal{Z}, \mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z})\rangle \\
&\geq \frac{p^2}{\sqrt{n_3}}\left(-\frac{1}{2}\|\mathcal{Z}\|_F + \langle\mathcal{Z}, \mathcal{P}_T(\mathcal{Z})\rangle\right) \\
&= \frac{p^2}{\sqrt{n_3}}\left(\frac{1}{\sqrt{n_3}}\langle\overline{\mathcal{Z}}, \overline{\mathcal{P}_T(\mathcal{Z})}\rangle - \frac{1}{2}\|\mathcal{Z}\|_F\right) \\
&= \frac{p^2}{n_3}\|\overline{\mathcal{P}_T(\mathcal{Z})}\|_F^2 - \frac{p^2}{2\sqrt{n_3}}\|\mathcal{Z}\|_F \\
&\geq (p^2 - \frac{p^2}{2\sqrt{n_3}})\|\mathcal{P}_T(\mathcal{Z})\|_F^2 - \frac{p^2}{2\sqrt{n_3}}\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_F^2
\end{aligned}$$

Then we have

$$\begin{aligned}
\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_F^2 &\geq p^2\frac{2\sqrt{n_3}-1}{2\sqrt{n_3}+p^2}\|\mathcal{P}_T(\mathcal{Z})\|_F^2 \\
&\geq \frac{1}{4n^2n_3^3}\|\mathcal{P}_T(\mathcal{Z})\|_F^2
\end{aligned}$$

It follows that

$$\begin{aligned}
&\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{TNN} = \|\overline{\mathcal{P}_{T^\perp}(\mathcal{Z})}\|_* \geq \|\overline{\mathcal{P}_{T^\perp}(\mathcal{Z})}\|_F \\
&\geq \sqrt{n_3}\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_F \geq \frac{1}{2nn_3}\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_F
\end{aligned}$$

which finishes the proof. ∎

Appendix C
Proof of Proposition IV.1 Condition 2

Three more lemmas will be introduced in this sections and their proofs are provided in the Appendix D.

We define $\ell_{\infty,2^*}$ norm for tensors to return the largest ℓ_{2^*} norm of the tensor row or tensor column of a third-order tensor.

$$\|\mathcal{Z}\|_{\infty,2*} := \max\left\{\max_i\sqrt{\sum_{b,k}\mathcal{Z}_{ibk}^2}, \max_j\sqrt{\sum_{a,k}\mathcal{Z}_{ajk}^2}\right\}$$

The following lemma states that $\mathcal{R}_\Omega(\mathcal{Z})$ is closed to $\mathcal{Z}$ in tensor spectral norm. The difference is bounded with ℓ_∞ norm and $\ell_{\infty,2^*}$ norm.

Lemma C.1: If p satisfies the condition in Theorem III.1, and $\mathcal{Z} \in \mathbb{R}^{n\times n\times n_3}$. Then for any constant $c > 0$, we have

$$\|\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{Z}\|_{op} \leq c\left(\frac{\log(nn_3)}{p}\|\mathcal{Z}\|_\infty + \sqrt{\frac{\log(nn_3)}{p}}\|\mathcal{Z}\|_{\infty,2^*}\right) \tag{51}$$

holds with probability at least $1 - (2nn_3)^{-(c-1)}$.

The lemma below bounds the $\ell_{\infty,2^*}$ distance between the terms $\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z})$ and $\mathcal{P}_T(\mathcal{Z})$.

Lemma C.2: If p satisfies the condition in Theorem III.1 for some c_2 sufficiently large, and $\mathcal{Z} \in \mathbb{R}^{n\times n\times n_3}$. Then

$$\|(\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{P}_T(\mathcal{Z})\|_{\infty,2^*} \leq \frac{1}{2}\|\mathcal{Z}\|_{\infty,2^*} + \frac{1}{2}\sqrt{\frac{n}{\mu_0 r}}\|\mathcal{Z}\|_\infty$$

with probability at least $1 - (2n^2n_3)^{-(c_2-1)}$.

Lemma C.3: If p satisfies the condition in Theorem III.1 for some c_3 sufficiently large, and $\mathcal{Z} \in \mathbb{R}^{n\times n\times n_3}$. Then

$$\|(\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T)\mathcal{Z}\|_\infty \leq \frac{1}{2}\|\mathcal{Z}\|_\infty \tag{52}$$

with probability at least $1 - 2n^{-(c_3-2)}n_3^{-(c_3-1)}$.

Proof of Proposition IV.1 Condition 2:

Proof: **(a)** We will first construct a tensor dual certificate $\mathcal{Y}$ and then show it satisfies both conditions here. We will use an approach called Golfing Scheme introduced by Gross [33] and we will follow the idea in [10] [22] where the strategy is to construct $\mathcal{Y}$ iteratively. Let Ω be a union of smaller sets Ω_t

such that $\Omega = \cup_{t=1}^{t_0}\Omega_t$ where $t_0 = 20\log(nn_3)$. For each t, we assume

$$\mathbb{P}\left[(i,j,k)\in\Omega_t\right] = q := 1-(1-p)^{1/t}$$

and it is easy to verify that it's equivalent to our original Ω. Define $\mathcal{R}_{\Omega_t}$ similarly to $\mathcal{R}_\Omega$ as follows

$$\mathcal{R}_{\Omega_t}(\mathcal{Z}) = \sum_{i,j,k}\frac{1}{q}\mathbf{1}_{(i,j,k)\in\Omega_t}\mathcal{Z}_{ijk}\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top$$

set $\mathcal{W}_0 = 0$ and for $t = 1, 2, \ldots, t_0$,

$$\mathcal{W}_t = \mathcal{W}_{t-1} + \mathcal{R}_{\Omega_t}\mathcal{P}_T\left(\mathcal{U}*\mathcal{V}^\top - \mathcal{P}_T(\mathcal{W}_{t-1})\right) \tag{53}$$

and tensor $\mathcal{Y} = \mathcal{W}_{t_0}$. By this construction we can see $\mathcal{P}_\Omega(\mathcal{Y}) = \mathcal{Y}$.

For $t = 0, 1, \ldots, t_0$, set $\mathcal{D}_t = \mathcal{U}*\mathcal{V}^\top - \mathcal{P}_T(\mathcal{W}_t)$. Then we have $\mathcal{D}_0 = \mathcal{U}*\mathcal{V}^\top$ and

$$\mathcal{D}_t = (\mathcal{P}_T - \mathcal{P}_T\mathcal{R}_{\Omega_t}\mathcal{P}_T)(\mathcal{D}_{t-1}) \tag{54}$$

Note that Ω_t is independent of $\mathcal{D}_t$, which implies

$$\|\mathcal{D}_t\|_F \le \|\mathcal{P}_T - \mathcal{P}_T\mathcal{R}_{\Omega_t}\mathcal{P}_T\|\,\|\mathcal{D}_{t-1}\|_F \le \frac{1}{2}\|\mathcal{D}_{t-1}\|_F$$

since $q \ge p/t_0 \ge c'\mu_0 r\log(nn_3)/n$, we have

$$\begin{aligned}
&\left\|\mathcal{P}_T(\mathcal{Y}) - \mathcal{U}*\mathcal{V}^\top\right\|_F = \|\mathcal{D}_{t_0}\|_F\\
&\le \left(\frac{1}{2}\right)^{t_0}\left\|\mathcal{U}*\mathcal{V}^\top\right\|_F \le \frac{1}{4(nn_3)^2}\sqrt{r} \le \frac{1}{4nn_3^2}
\end{aligned}$$

holds with probability at least $1 - c'(2nn_3)^{-c''}$ by the union bound, for some large enough constants $c', c'' > 0$.

(b) From (53) we know that $\mathcal{Y} = \mathcal{W}_{t_0} = \sum_{t=1}^{t_0}(\mathcal{R}_{\Omega_t}\mathcal{P}_T)(\mathcal{D}_{t-1})$, so use **Lemma** C.1 we obtain for some constant $c > 0$,

$$\begin{aligned}
&\|\mathcal{P}_{T^\perp}(\mathcal{Y})\|_{op} \le \sum_{t=1}^{t_0}\|\mathcal{P}_{T^\perp}\left(\mathcal{R}_{\Omega_t}\mathcal{P}_T\right)(\mathcal{D}_{t-1})\|_{op}\\
&\le \sum_{t=1}^{t_0}\|(\mathcal{R}_{\Omega_t} - \mathcal{I})\,\mathcal{P}_T(\mathcal{D}_{t-1})\|_{op}\\
&\le c\sum_{t=1}^{t_0}\left(\frac{\log(nn_3)}{q}\|\mathcal{D}_{t-1}\|_\infty + \sqrt{\frac{\log(nn_3)}{q}}\|\mathcal{D}_{t-1}\|_{\infty,2^*}\right)\\
&\le \frac{c}{\sqrt{c_0}}\sum_{t=1}^{t_0}\left(\frac{n}{\mu_0 r}\|\mathcal{D}_{t-1}\|_\infty + \sqrt{\frac{n}{\mu_0 r}}\|\mathcal{D}_{t-1}\|_{\infty,2^*}\right)
\end{aligned}$$

where we could bound term $\|\mathcal{D}_{t-1}\|_\infty$ using **Lemma** C.3 as follows,

$$\begin{aligned}
&\|\mathcal{D}_{t-1}\|_\infty\\
&= \|(\mathcal{P}_T - \mathcal{P}_T\mathcal{R}_{\Omega_{t-1}}\mathcal{P}_T)\ldots(\mathcal{P}_T - \mathcal{P}_T\mathcal{R}_{\Omega_1}\mathcal{P}_T)(\mathcal{D}_0)\|_\infty\\
&\le \left(\frac{1}{2}\right)^{t-1}\|\mathcal{U}*\mathcal{V}^\top\|_\infty
\end{aligned} \tag{55}$$

and $\|\mathcal{D}_{t-1}\|_{\infty,2^*}$ is bounded using **Lemma** C.2 and (54) (55),

$$\begin{aligned}
&\|\mathcal{D}_{t-1}\|_{\infty,2^*} = \|(\mathcal{P}_T - \mathcal{P}_T\mathcal{R}_{\Omega_{k-1}}\mathcal{P}_T)(\mathcal{D}_{t-2})\|_{\infty,2^*}\\
&\le \frac{1}{2}\|\mathcal{D}_{t-2}\|_{\infty,2^*} + \frac{1}{2}\sqrt{\frac{n}{\mu_0 r}}\|\mathcal{D}_{t-2}\|_\infty\\
&\le \frac{1}{2}\left(\frac{1}{2}\|\mathcal{D}_{t-3}\|_{\infty,2^*} + \frac{1}{2}\sqrt{\frac{n}{\mu_0 r}}\|\mathcal{D}_{t-3}\|_\infty\right)\\
&\quad + \frac{1}{2}\sqrt{\frac{n}{\mu_0 r}}\|\mathcal{D}_{t-2}\|_\infty\\
&\le t\left(\frac{1}{2}\right)^{t-1}\sqrt{\frac{n}{\mu_0 r}}\left\|\mathcal{U}*\mathcal{V}^\top\right\|_\infty + \left(\frac{1}{2}\right)^{t-1}\left\|\mathcal{U}*\mathcal{V}^\top\right\|_{\infty,2^*}
\end{aligned}$$

So we get

$$\begin{aligned}
&\|\mathcal{P}_{T^\perp}(\mathcal{Y})\|_{op}\\
&\le \frac{c}{\sqrt{c_0}}\frac{n}{\mu_0 r}\|\mathcal{U}*\mathcal{V}^\top\|_\infty\sum_{t=1}^{t_0}(t+1)\left(\frac{1}{2}\right)^{t-1}\\
&\quad + \frac{c}{\sqrt{c_0}}\sqrt{\frac{n}{\mu_0 r}}\|\mathcal{U}*\mathcal{V}^\top\|_{\infty,2^*}\sum_{t=1}^{t_0}\left(\frac{1}{2}\right)^{t-1}\\
&\le \frac{6c}{\sqrt{c_0}}\frac{n}{\mu_0 r}\|\mathcal{U}*\mathcal{V}^\top\|_\infty + \frac{2c}{\sqrt{c_0}}\sqrt{\frac{n}{\mu_0 r}}\left\|\mathcal{U}*\mathcal{V}^\top\right\|_{\infty,2^*}
\end{aligned}$$

holds with probability at least $1 - c'(2nn_3)^{-c''}$ by the union bound for some large enough constants $c', c'' > 0$.

Now let's bound $\|\mathcal{U}*\mathcal{V}^\top\|_\infty$. We have

$$\begin{aligned}
\|\mathcal{U}*\mathcal{V}^\top\|_\infty &= \max_{i,j,k}\left(\mathcal{U}(i,:,:)*\mathcal{V}^\top(:,j,:)\right)_k\\
&= \max_{i,j}\|\mathcal{U}(i,:,:)*\mathcal{V}^\top(:,j,:)\|_\infty
\end{aligned}$$

Note the fact that for two tensor tubes $\mathring{\boldsymbol{x}}, \mathring{\boldsymbol{y}} \in \mathbb{R}^{1\times1\times n_3}$, use the Cauchy-Schwartz inequality we get

$$\|\mathring{\boldsymbol{x}} * \mathring{\boldsymbol{y}}\|_\infty \le \|\mathring{\boldsymbol{x}}\|_{2^*}\|\mathring{\boldsymbol{y}}\|_{2^*}$$

Then let $\mathring{\boldsymbol{u}}_t = \mathcal{U}(i,t,:), \mathring{\boldsymbol{v}}_t^\top = \mathcal{V}^\top(t,j,:)$, we can further write $\|\mathcal{U}*\mathcal{V}^\top\|_\infty$ as follows

$$\begin{aligned}
\|\mathcal{U}*\mathcal{V}^\top\|_\infty &= \max_{i,j}\left\|\sum_{t=1}^{r}\mathring{\boldsymbol{u}}_t * \mathring{\boldsymbol{v}}_t^\top\right\|_\infty\\
&\le \max_{i,j}\sum_{t=1}^{r}\|\mathring{\boldsymbol{u}}_t * \mathring{\boldsymbol{v}}_t^\top\|_\infty \le \max_{i,j}\sum_{t=1}^{r}\|\mathring{\boldsymbol{u}}_t\|_{2^*}\|\mathring{\boldsymbol{v}}_t^\top\|_{2^*}\\
&\le \max_{i,j}\sum_{t=1}^{r}\frac{1}{2}\left(\|\mathring{\boldsymbol{u}}_t\|_{2^*}^2 + \|\mathring{\boldsymbol{v}}_t^\top\|_{2^*}^2\right)\\
&= \max_{i,j}\left\{\frac{1}{2}\|\vec{e}_i^\top * \mathcal{U}\|_{2^*}^2 + \frac{1}{2}\|\mathcal{V}^\top * \vec{e}_j\|_{2^*}^2\right\}\\
&\le \frac{\mu_0 r}{n}
\end{aligned}$$

by the standard incoherence condition. We also have

$$\|\mathcal{U} * \mathcal{V}^\top\|_{\infty,2^*} = \max_{i,j} \left\{ \|\mathcal{U} * \mathcal{V}^\top * \vec{e}_i\|_{2^*}, \|\vec{e}_j^\top * \mathcal{U} * \mathcal{V}^\top\|_{2^*} \right\} \leq \sqrt{\frac{\mu_0 r}{n}}$$

and thus

$$\|\mathcal{P}_{T^\perp}(\mathcal{Y})\| \leq \frac{8c}{\sqrt{c_0}} \leq \frac{1}{2}$$

given c_0 large enough. ■

Appendix D
Proofs of Supporting Lemmas

Proof of Lemma C.1:

Proof: Let

$$\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{Z} = \sum_{i,j,k} \mathcal{C}_{(ijk)} = \sum_{i,j,k} \left(\frac{1}{p}\delta_{ijk} - 1\right) \mathcal{Z}_{ijk} \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top$$

where $\mathcal{C}_{(ijk)}$ are independent tensors. Then we have

$$\overline{\mathcal{C}_{(ijk)}} = \sum_{i,j,k} \left(\frac{1}{p}\delta_{ijk} - 1\right) \mathcal{Z}_{ijk} \overline{\vec{e}_i}\, \overline{\mathring{e}_k}\, \overline{\vec{e}_j}^\top$$

Notice that $\mathbb{E}\left[\overline{\mathcal{C}_{(ijk)}}\right] = 0$ and $\|\overline{\mathcal{C}_{(ijk)}}\| \leq \frac{1}{p}\|\mathcal{Z}\|_\infty$. Moreover,

$$\left\| \mathbb{E}\left[\sum_{i,j,k} \overline{\mathcal{C}_{(ijk)}}^\top \overline{\mathcal{C}_{(ijk)}}\right] \right\| = \left\| \mathbb{E}\left[\sum_{i,j,k} \mathcal{C}_{(ijk)}^\top \mathcal{C}_{(ijk)}\right] \right\|_{op}$$
$$= \left\| \sum_{i,j,k} \mathcal{Z}_{ijk}^2 \vec{e}_j * \vec{e}_j^\top \mathbb{E}\left(\frac{1}{p}\delta_{ijk} - 1\right)^2 \right\|_{op}$$
$$= \left\| \frac{1-p}{p} \sum_{i,j,k} \mathcal{Z}_{ijk}^2 \vec{e}_j * \vec{e}_j^\top \right\|_{op}$$

since $\vec{e}_j * \vec{e}_j^\top$ will return a zero tensor except for $(j,j,1)$th entry equaling 1, we have

$$\left\| \mathbb{E}\left[\sum_{i,j,k} \overline{\mathcal{C}_{(ijk)}}^\top \overline{\mathcal{C}_{(ijk)}}\right] \right\| = \frac{1-p}{p} \max_j \left| \sum_{i,k} \mathcal{Z}_{ijk} \right| \leq \frac{1}{p}\|\mathcal{Z}\|_{\infty,2^*}^2$$

And $\left\|\mathbb{E}\left[\sum_{i,j,k} \overline{\mathcal{C}_{(ijk)}}\, \overline{\mathcal{C}_{(ijk)}}^\top\right]\right\|$ is bounded similarly. Then use the extension of **Theorem A.1**, for any $c' > 0$ we have

$$\|\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{Z}\|_{op} = \|\overline{\mathcal{R}_\Omega(\mathcal{Z})} - \overline{\mathcal{Z}}\| = \left\| \sum_{i,j,k} \overline{\mathcal{C}_{(ijk)}} \right\|$$
$$\leq \sqrt{\frac{4c'}{p}\|\mathcal{Z}\|_{\infty,2^*}^2 \log(2nn_3)} + \frac{c'}{p}\|\mathcal{Z}\|_\infty \log(2nn_3)$$
$$\leq c\left(\frac{\log(nn_3)}{p}\|\mathcal{Z}\|_\infty + \sqrt{\frac{\log(nn_3)}{p}}\|\mathcal{Z}\|_{\infty,2^*}\right)$$

holds with probability at lease $1-(2nn_3)^{-(c-1)}$ for any $c \geq \max\{c', 2\sqrt{c'}\}$.

Proof of Lemma C.2:

Proof: Consider any bth tensor column of $\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{P}_T(\mathcal{Z})$:

$$(\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{P}_T(\mathcal{Z})) * \vec{e}_b = \sum_{i,j,k}(\frac{1}{p}\delta_{ijk} - 1)\mathcal{Z}_{ijk}\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top) * \vec{e}_b = \sum_{i,j,k} \vec{a}_{ijk}$$

where $\vec{a}_{ijk} \in \mathbb{R}^{n\times 1\times n_3}$ are zero-mean independent tensor columns. Let $\tilde{a}_{ijk} \in \mathbb{R}^{nn_3\times 1}$ be the vectorized column vector of $\vec{a}_{ijk}$. Then the ℓ_2 norm of the vector $\tilde{a}_{ijk}$ is bounded by the following

$$\|\tilde{a}_{ijk}\| = \|\vec{a}_{ijk}\|_{2^*} \leq \frac{1-p}{p}\mathcal{Z}_{ijk}\left\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top) * \vec{e}_b\right\|_{2^*}$$
$$\leq \frac{1}{p}\sqrt{\frac{2\mu_0 r}{n}}\|\mathcal{Z}\|_\infty \leq \frac{1}{c_0\log(nn_3)}\sqrt{\frac{2n}{\mu_0 r}}\|\mathcal{Z}\|_\infty$$

for some constant $c_0 > 0$ given p satisfying (26). We also have

$$\left|\mathbb{E}\left[\sum_{i,j,k}\tilde{a}_{ijk}^\top\tilde{a}_{ijk}\right]\right| = \mathbb{E}\left[\sum_{i,j,k}\|\vec{a}_{ijk}\|_{2^*}^2\right] = \frac{1-p}{p}\sum_{i,j,k}\mathcal{Z}_{ijk}^2\left\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top) * \vec{e}_b\right\|_{2^*}^2$$

Use the definition of $\mathcal{P}_T$ and the incoherent condition, we can write

$$\left\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top) * \vec{e}_b\right\|_{2^*}$$
$$= \Big\|(\mathcal{U} * \mathcal{U}^\top * \vec{e}_i * \mathring{e}_k) * \mathring{e}_j^\top * \vec{e}_b + (\mathcal{I} - \mathcal{U} * \mathcal{U}^\top) * \vec{e}_i * \mathring{e}_k * \mathring{e}_j^\top * \mathcal{V} * \mathcal{V}^\top * \vec{e}_b\Big\|_{2^*}$$
$$\leq \sqrt{\frac{\mu_0 r}{n}}\left\|\vec{e}_j^\top * \vec{e}_b\right\|_{2^*} + \left\|(\mathcal{I} - \mathcal{U} * \mathcal{U}^\top) * \vec{e}_i * \vec{e}_k\right\|\left\|\vec{e}_j^\top * \mathcal{V} * \mathcal{V}^\top * \vec{e}_b\right\|_{2^*}$$
$$\leq \sqrt{\frac{\mu_0 r}{n}}\left\|\vec{e}_j^\top * \vec{e}_b\right\|_{2^*} + \left\|\vec{e}_j^\top * \mathcal{V} * \mathcal{V}^\top * \vec{e}_b\right\|_{2^*}$$

where $\mathfrak{I}$ is the identity tensor. Thus,

$$\begin{aligned}
&\left|\mathbb{E}\left[\sum_{i,j,k} \tilde{a}_{ijk}^\top \tilde{a}_{ijk}\right]\right| \\
&\leq \frac{2}{p}\sum_{ijk} \mathcal{Z}_{ijk}^2 \frac{\mu_0 r}{n}\left\|\vec{e}_j^\top * \vec{e}_b\right\|_{2^*}^2 \\
&\quad + \frac{2}{p}\sum_{ijk} \mathcal{Z}_{ijk}^2 \left\|\vec{e}_j^\top * \mathcal{V} * \mathcal{V}^\top * \vec{e}_b\right\|_{2^*}^2 \\
&= \frac{2\mu_0 r}{pn}\sum_{i,k}\mathcal{Z}_{ibk}^2 + \frac{2}{p}\sum_j \left\|\vec{e}_j^\top * \mathcal{V} * \mathcal{V}^\top * \vec{e}_b\right\|_{2^*}^2 \sum_{i,k}\mathcal{Z}_{ijk}^2 \\
&\leq \frac{2\mu_0 r}{pn}\|\mathcal{Z}\|_{\infty,2^*}^2 + \frac{2}{p}\left\|\mathcal{V} * \mathcal{V}^\top * \vec{e}_b\right\|_{2^*}^2 \|\mathcal{Z}\|_{\infty,2^*}^2 \\
&\leq \frac{4\mu_0 r}{pn}\|\mathcal{Z}\|_{\infty,2^*}^2 \leq \frac{4}{c_0\log(nn_3)} \qquad (56)
\end{aligned}$$

where (56) is because $\vec{e}_j^\top * \vec{e}_b = 0$ if $j \neq b$. In the same fashion $|\mathbb{E}[\sum_{i,j,k}\tilde{a}_{ijk}\tilde{a}_{ijk}^\top]|$ is bounded by the exact same quantity. Since the spectral norm of the vector $\tilde{a}_{ijk}$ is equal to its ℓ_2 norm, then use the extension of **Theorem** A.1 we have for any $c_1 > 0$, we have

$$\begin{aligned}
&\|(\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{P}_T(\mathcal{Z})) * \vec{e}_b\|_{2^*} \\
&= \left\|\sum_{i,j,k}\vec{a}_{ijk}\right\|_{2^*} = \left\|\sum_{i,j,k}\tilde{a}_{ijk}\right\| \\
&\leq \sqrt{4c_1\sigma^2\log(nn_3)} + c_1 M\log(nn_3) \\
&\leq \frac{1}{2}\|\mathcal{Z}\|_{\infty,2^*} + \frac{1}{2}\sqrt{\frac{n}{\mu_0 r}}\|\mathcal{Z}\|_\infty
\end{aligned}$$

holds with probability at least $1 - (nn_3)^{-(c_2-1)}$ for c_2 large enough.

We can also do the same to the tensor rows $\vec{e}_a^\top * (\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{P}_T(\mathcal{Z}))$ and get the same bound. Then using a union bound over all the tensor columns and tensor rows, the result holds with probability at least $1 - 2n^2 n_3^{-(c_2-1)}$. With c_2 large enough, the probability goes to zero. Done.

Proof of Lemma C.3:

Proof: Observe that

$$\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z}) = \sum_{i,j,k}\frac{1}{p}\delta_{ijk}\mathcal{Z}_{ijk}\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)$$

so we have that any (a,b,c)th entry of $\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z}) - \mathcal{P}_T(\mathcal{Z})$ is given by

$$\begin{aligned}
&\langle\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z}) - \mathcal{P}_T(\mathcal{Z}), \vec{e}_a * \mathring{e}_c * \vec{e}_b^\top\rangle \\
&= \sum_{i,j,k}\left(\frac{\delta_{ijk}}{p} - 1\right)\mathcal{Z}_{ijk}\langle\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top), \vec{e}_a * \mathring{e}_c * \vec{e}_b^\top\rangle \\
&:= \sum_{i,j,k}\mathcal{H}_{ijk,abc}
\end{aligned}$$

It is easy to observe that

$$\begin{aligned}
&|\mathcal{H}_{ijk,abc}| \\
&\leq \frac{1}{p}\|\mathcal{Z}\|_\infty\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\|_F\|\mathcal{P}_T(\vec{e}_a * \mathring{e}_c * \vec{e}_b^\top)\|_F \\
&\leq \frac{2\mu_0 r}{np}\|\mathcal{Z}\|_\infty
\end{aligned}$$

We also have

$$\begin{aligned}
&\left|\mathbb{E}\left[\sum_{i,j,k}\mathcal{H}_{ijk,abc}^2\right]\right| \\
&= \frac{1-p}{p}\|\mathcal{Z}\|_\infty^2\sum_{i,j,k}\left|\langle\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top), \vec{e}_a * \mathring{e}_c * \vec{e}_b^\top\rangle\right|^2 \\
&= \frac{1-p}{p}\|\mathcal{Z}\|_\infty^2\left\|\mathcal{P}_T(\vec{e}_a * \mathring{e}_c * \vec{e}_b^\top)\right\|_F \leq \frac{2\mu_0 r}{np}\|\mathcal{Z}\|_\infty^2
\end{aligned}$$

Then use **Theorem A.1**, we have

$$\begin{aligned}
&\mathbb{P}\left[(\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z}) - \mathcal{P}_T(\mathcal{Z}))_{abc} \geq \frac{1}{2}\|\mathcal{Z}\|_\infty\right] \\
&\leq 2\exp\left(\frac{-\|\mathcal{Z}\|_\infty^2/4}{\frac{2\mu_0 r}{np}\|\mathcal{Z}\|_\infty^2 + \frac{\mu_0 r}{3np}\|\mathcal{Z}\|_\infty^2}\right) \\
&\leq 2(nn_3)^{-c_3}
\end{aligned}$$

for some $c_3 = 3c_0/28$ large enough, given p satisfying (24). Then using the union bound on every (a,b,c)th entry we have $\|(\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T)(\mathcal{Z})\|_\infty \leq \frac{1}{2}\|\mathcal{Z}\|_\infty$ holds with probability at least $1 - 2n^{-(c_3-2)}n_3^{-(c_3-1)}$.

REFERENCES

[1] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus (Springer Series in Computational Mathematics)*. New York, NY, USA: Springer-Verlag, 2012, vol. 42.

[2] N. Boumal and P.-A. Absil, "RTRMC: A Riemannian trust-region method for low-rank matrix completion," in *Proc. Adv Neural Inf Process. Sys. 24*, 2011, pp. 406–414.

[3] N. Linial, E. London, and Y. Rabinovich, "The geometry of graphs and some of its algorithmic applications," *Combinatorica*, vol. 15, no. 2, pp. 215–245, 1995.

[4] G. Obozinski, B. Taskar, and M. I. Jordan, "Joint covariate selection and joint subspace selection for multiple classification problems," *Stat. Comput.*, vol. 20, no. 2, pp. 231–252, 2010.

[5] T. Kolda and J. Sun, "Scalable tensor decompositions for multi-aspect data mining," in *Proc. 8th IEEE Int. Conf. Data Mining*, Dec. 2008, pp. 363–372.

[6] J. Sun, S. Papadimitriou, C. Lin, N. Cao, S. Liu, and W. Qian, "Multivis: Content-based social network exploration through multi-way visual analysis," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 30–May 2 2009, Sparks, NV, USA, 2009, pp. 1064–1075.

[7] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

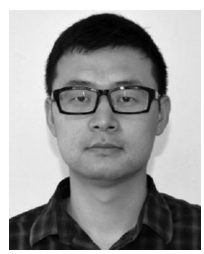
[8] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3842–3849.

[9] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, 2012.

[10] B. Recht, "A simpler approach to matrix completion," *J. Mach. Learn. Res.*, vol. 12, pp. 3413–3430, 2011.

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

[12] V. D. Silva and L.-H. Kim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1084–1127, 2008.
[13] L. Grasedyck, "Hierarchical singular value decomposition of tensors," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 4, pp. 2029–2054, May 2010.
[14] B. Huang, C. Mu, D. Goldfarb, and J. Wright, "Provable low-rank tensor recovery," *Optim.-Online*, p. 4252, 2014.
[15] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Problems*, vol. 27, no. 2, 2011, Art. no. 025010.
[16] K. Braman, "Third-order tensors as linear operators on a space of matrices," *Linear Algebra Appl.*, vol. 433, no. 7, pp. 1241–1253, 2010.
[17] M. Kilmer, K. Braman, N. Hao, and R. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.
[18] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, 2011.
[19] D. F. Gleich, C. Greif, and J. M. Varah, "The power and Arnoldi methods in an algebra of circulants," *Numer. Linear Algebra Appl.*, vol. 20, no. 5, pp. 809–831, 2013.
[20] C. Navasca, M. Opperman, T. Penderghest, and C. Tamon, "Tensors as module homomorphisms over group rings," arXiv 1005.1894, May 2010.
[21] E. Kernfeld, M. Kilmer, and S. Aeron, "Tensor-tensor products with invertible linear transforms," *Linear Algebra Appl.*, vol. 485, pp. 545–570, 2015.
[22] Y. Chen, "Incoherence-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2909–2923, May 2015.
[23] C. Mu, B. Huang, J. Wright, and D. Goldfarb, "Square deal: Lower bounds and improved relaxations for tensor recovery," in *Proc. 31th Int. Conf. Mach. Learn.*, Beijing, China, 21–26 June 2014, pp. 73–81.
[24] P. Jain and S. Oh, "Provable tensor factorization with missing data," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Eds. Sydney, NSW, Australia: Curran Associates, 2014, pp. 1431–1439.
[25] A. Krishnamurthy and A. Singh, "Low-rank matrix and tensor completion via adaptive sampling," in *Proc. Neural Inf. Process. Syst.*, 2013, pp. 836–844.
[26] L. Karlsson, D. Kressner, and A. Uschmajew, "Parallel algorithms for tensor completion in the CP format," *Parallel Comput.*, vol. 57, pp. 222–234, Sep. 2016.
[27] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.
[28] B. Romera-Paredes and M. Pontil, "A new convex relaxation for tensor completion," in *Proc. Adv. Neural Inf. Process. Syst. 26*, Lake Tahoe, NV, USA, 2013, pp. 2967–2975.
[29] E. Kernfeld, S. Aeron, and M. E. Kilmer, "Clustering multi-way data: A novel algebraic approach," vol. abs/arXiv 1412.7056, 2014.
[30] O. Semerci, N. Hao, M. E. Kilmer, and E. L. Miller, "Tensor-based formulation and nuclear norm regularization for multi-energy computed tomography," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1678–1693, Apr. 2014.
[31] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.
[32] D. Gross and V. Nesme, "Note on sampling without replacing from a finite collection of matrices," arXiv 1001.2738, 2010.
[33] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1548–1566, Mar. 2011.
[34] X.-Y. Liu, V. Aggarwal, X. Wang, S. Aeron, and M.-Y. Wu, "Adaptive sampling of RF fingerprints for fine-grained indoor localization," *IEEE Trans. Mobile Comput.*, vol. 15, no. 10, pp. 2411–2423, Oct. 2016.
[35] G. Watson, "Characterization of the subdifferential of some matrix norms," *Linear Algebra Appl.*, vol. 170, pp. 33–45, 1992.
[36] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Mar. 2010.
[37] D. Kressner, M. Steinlechner, and B. Vandereycken, "Low-rank tensor completion by riemannian optimization," *Bit Numer. Math.*, vol. 54, pp. 447–468, 2014.
[38] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 569–579, Mar. 2002.
[39] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Found. Comput. Math.*, vol. 12, no. 4, pp. 389–434, Aug. 2012.



Zemin Zhang (S'14) received the B.Sc. degree in electronics information engineering from Tsinghua University, Beijing, China, in 2011. He is currently working toward the Ph.D. degree in electrical engineering and computer science at Tufts University, Medford, MA, USA. His research interests include computer vision, machine learning, and tensor analysis and applications.

Shuchin Aeron (SM'05) received the M.S. and Ph.D. degrees from Boston University, Boston, MA, USA, in 2004 and 2009, respectively. He is currently an Assistant Professor in the department of ECE, Tufts University, Medford, MA , USA. Following that he was a Postdoctoral Research Fellow at Schlumberger-Doll Research, Cambridge, MA, USA. His research interests include information theory and tensor algebraic methods for statistical signal processing with applications to computer vision and geophysical signal processing. He received the 2016 NSF CAREER award.