Applied Mathematical Modelling 138 (2025) 115744

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Adaptive sampling with tensor leverage scores for exact low-rank third-order tensor completion

Xuan Chen [a,b,c], Tai-Xiang Jiang [a,b,c,*], Yexun Hu [a,b,c], Jinjin Yu [a,b,c], Michael K. Ng [d]

[a] School of Computing and Artificial Intelligence, Southwestern University of Finance and Economics, Chengdu, People's Republic of China
[b] Kash Institute of Electronics and Information Industry, Kashi, People's Republic of China
[c] Engineering Research Center of Intelligent Finance, Ministry of Education, Chengdu, People's Republic of China
[d] Department of Mathematics, Hong Kong Baptist University, Hong Kong

ARTICLE INFO

MSC:
15A69
65F99

Keywords:
Tensor completion
Adaptive sampling
Tensor singular value decomposition
Low-rank

ABSTRACT

Tensor completion aims at estimating the missing entries from the incomplete observation. Under the tensor singular value decomposition framework, the exact recovery of a low-tubal-rank third-order tensor could be achieved via convex optimization with high probability if the tensor satisfies the tensor incoherence condition. In this work, we show that, when the random selection of entries is made adaptive to a distribution which is dependent on the coherence structure of the tensor, any low-tubal-rank tensor of the size $n \times n \times n$ with tubal-rank r can be exactly recovered with high probability from as few as $O(rn^2 \log^2(n))$ randomly chosen entries. In practice, tensor leverage scores are not known a priori, and we design a two-phase adaptive sampling strategy to obtain the leverage scores. Numerical experiments on synthetic and real-world third-order tensor data sets are used to validate our theoretical results and illustrate that the tensor recovery performance of the proposed two-phase adaptive sampling scheme is better than that of the other state-of-the-art methods.

1. Introduction

The main aim of tensor completion is to estimate the missing or unobserved entries from incomplete observation. It is an important problem that has appeared in the literature of a diverse set of fields including color image inpainting [1–4], video inpainting [5], and remote sensing hyperspectral image restoration [6,7] to name a few. Similar to matrix completion, tensor completion is an ill-posed problem and some prior or assumptions are needed. For the matrix case, the common assumption is to suppose that the unknown matrix is low-rank or approximately low-rank. One can recover a low-rank matrix by solving the following convex optimization problem

$$\min_{\mathbf{X}\in\mathbb{R}^{n_1\times n_2}} \quad \|\mathbf{X}\|_* \quad \text{s.t. } \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{O}), \tag{1}$$

where $\mathbf{X}$ is the underlying matrix, $\mathbf{O} \in \mathbb{R}^{n_1\times n_2}$ is the original matrix, Ω is the support of the observed entries, $\mathcal{P}_\Omega(\cdot)$ is the projection operator keeping the elements in Ω while making others to be zeros, and $\|\cdot\|_*$ is the matrix nuclear norm defined as the sum of

* Corresponding author.
E-mail addresses: 41908140@smail.swufe.edu.cn (X. Chen), taixiangjiang@gmail.com (T.-X. Jiang), 220081202001@smail.swufe.edu.cn (Y. Hu), 223081200054@smail.swufe.edu.cn (J. Yu), michael-ng@hkbu.edu.hk (M.K. Ng).

https://doi.org/10.1016/j.apm.2024.115744
Received 2 April 2024; Received in revised form 29 September 2024; Accepted 1 October 2024

Available online 10 October 2024

singular values, being the convex envelope of the matrix rank within a certain set. The nuclear norm of $\mathbf{A}$ can also be calculated as $\text{trace}(\sqrt{\mathbf{A}^*\mathbf{A}})$, where $*$ means complex conjugate.

The low-rank matrix completion (LRMC) problem in (1) can be efficiently optimized by the alternating direction method of multipliers (ADMM) [8,9] with the singular value thresholding (SVT) operator [10]. Candès and Recht [11] showed that one can perfectly recover most low-rank matrices from an incomplete set of entries, which are uniformly sampled at random, via (1) with high probability (w.h.p.) if $\mathbf{O}$ satisfies the incoherence condition. Intuitively, the incoherence condition requires the matrix $\mathbf{O}$ not to be sparse. Also, when the incoherence condition is satisfied, nuclear norm minimization can produce the minimum rank solution [12]. Candès and Tao [13] presented the information limit for the minimum number of entries needed to recover a matrix of rank r exactly by any method in follow-on work. Then, Recht [14] provided a simpler proof with the help of the Noncommutative Bernstein Inequality [15]. In [16], the joint incoherence condition is successfully removed by using the $\ell_{\infty,2}$ norm to get a similar bound when constructing a dual certificate to ensure the solution is the unique minimizer of the optimization program. In [17], Chen et al. showed that the incoherence is actually required because of the uniform sampling and further considers the leveraged sampling, i.e., random sampling conducted according to a specific biased distribution dependent on the coherence structure of the matrix.

When it comes to tensors, three categories of low-rank tensor completion (LRTC) methods can be found.[1] They are based on the CANDECOMP/PARAFAC (CP) decomposition [21–23], the Tucker decomposition [24,2], and the tensor train (TT) or tensor ring (TC) decomposition [25–28], respectively. As discussed in [29], being NP-hard to compute the CP-rank [18] and unrevealed convex envelope of the CP-rank make the low-CP-rank tensor recovery challenging. Meanwhile, provable guarantees for low-Tucker-rank tensor recovery [30–32] and low TT (or TR) -rank tensor recovery [33,34] are given with employing the theory and methods for matrix completion after *matricizing* the tensor in various ways.

In this work, we mainly focus on another type of LRTC methods based on the tensor singular value decomposition (t-SVD) framework [35–38], in which the fundamental tensor-tensor product (t-prod) operation is closed on the set of third-order tensors and allows new extensions of familiar matrix analysis to the third-order tensors while avoiding the loss of information inherent in *matricizing* or *flattening* [38]. More specifically, we consider the following tensor tubal nuclear norm (TNN, subsequently defined in Definition 10) minimization model [39] for third-order tensor completion:

$$\min_{\mathcal{X}\in\mathbb{R}^{n_1\times n_2\times n_3}} \quad \|\mathcal{X}\|_{\text{TNN}} \quad \text{s.t. } \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{O}), \tag{2}$$

where $\mathcal{X}$ is the underlying tensor, $\mathcal{O}\in\mathbb{R}^{n_1\times n_2\times n_3}$ is the original tensor. As shown in [39], a tensor $\mathcal{O}$ of the tubal-rank r can be exactly recovered w.h.p. via solving (2) provided that elements are uniformly sampled at random and $\mathcal{O}$ satisfies the tensor incoherent condition. Similar to the matrix case, required number of measurements in (2) is in direct proportion to $rn_{(1)}n_3\log^2(n_{(1)}n_3)$, where $n_{(1)} = \max(n_1, n_2)$. In [29], linear invertible transform is used to construct the t-prod and t-SVD framework. Song et al. [40] theoretically provide bounds for exact robust tensor completion using unitary transform based t-SVD. In [41], Liu et al. also minimized the TNN to complete the fingerprint data, while they further consider a two-pass adaptive tubal-wise sampling strategy in their indoor localization application. Provided that the total sampling budget is M tubes, in the 1st-pass sampling, $\sigma M(0<\sigma<1)$ tubes are sampled uniformly at random. Then, in the 2nd-pass sampling, remaining $(1-\sigma)M$ tubes are allocated to those highly informative columns identified by the 1st-pass sampling. This two-pass adaptive sampling strategy results in an attractive improvement over (uniform) random sampling in localization accuracy for the same sample complexity. Meanwhile, completing the coherent low-tubal-rank tensor is also studied in a pioneer work [42], while the authors only studied the sufficient condition of exact completion without experimental validations.

The main aim of this work is to break through the requirement of the tensor incoherence via providing the theoretical guarantees for any low-rank tensor completion from a subset of entries, which are adaptively sampled at random. Our main contributions are listed as follows:

- The first contribution which is also our key innovation is that we define tensor leverage scores, which can be viewed as the local versions of the standard tensor incoherence parameter to measure the coherence structure of the low-rank tensor. Then, we prove that any low-tubal-rank tensor (of the size $n\times n\times n$ and tubal-rank r) can be exactly recovered w.h.p. from as few as $O(rn^2\log^2(n))$ observed entries via solving the convex optimization problem in (2), if the probability of an element being observed is in proportion to the sum of related tensor leverage scores.
- A two-phase adaptive sampling strategy is provided when tensor leverage scores are not known a priori. When the amount of elements to be sampled is determined, a portion of entries is first sampled uniformly at random. Then, the tensor leverage scores are estimated via t-SVD on the samples in the first phase. Then, the remaining entries are sampled according to estimated tensor leverage scores.
- Numerical results on the synthetic data are reported to verify our theoretical results and the effectiveness of the two-phase adaptive sampling. Experiments on color images further illustrate the superiority of our method over state-of-the-art methods. Meanwhile, for the traffic data imputation, it is shown that the temporal periodicity can be utilized to well estimate the leverage score, further illustrating the practical application value of our method.

[1] Please refer to [18–20] for a comprehensive overview.

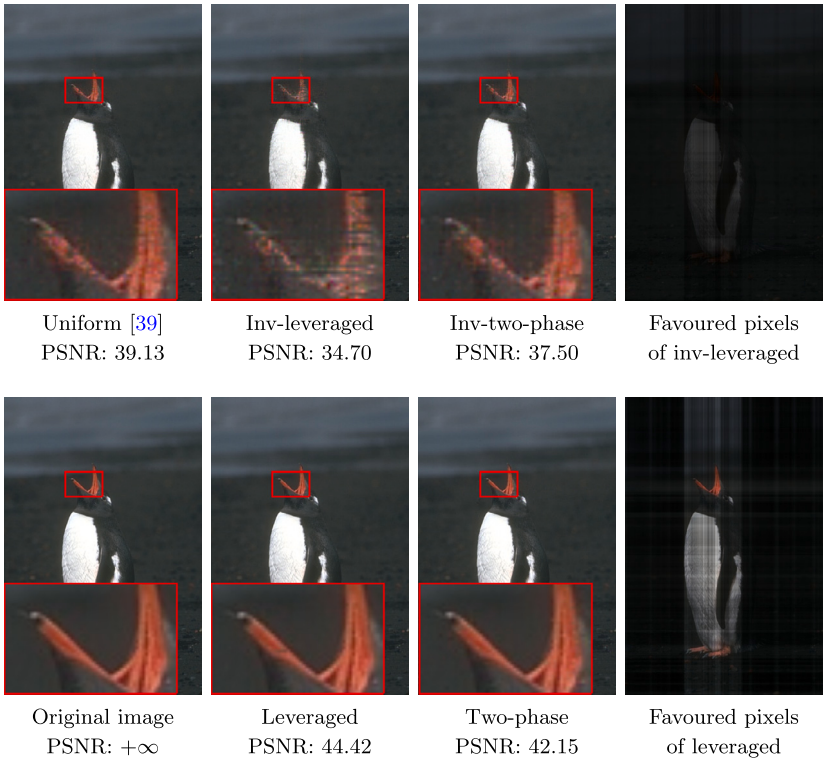

Fig. 1. Tensor completion results for a color image with 50% missing entries with different sampling strategies. All the results are obtained via minimizing the tensor nuclear norm [39]. "Uniform" indicates uniform sampling at random. "Leveraged" denotes adaptive sampling according to probabilities in direct proportion to tensor leverage scores while probabilities for "Inv-leveraged" are *inversely* proportional to tensor leverage scores. "Two-phase" and "Inv-two-phase" stand for random sampling in the second phase using the leveraged sampling and inv-leveraged sampling, respectively. The favored pixels in the last column are from the element-wise multiplication between the probabilities, which are scaled to $[0, 1]$, and the original image.

We show in Fig. 1 that when the random sampling of a real-world color image is conducted according to the tensor leverage scores, i.e., "Leveraged" and "Two-phase" (to be introduced in Sect. 3.2), the color image completion results are obviously better than that of the uniformly random sampling. Moreover, if elements are randomly sampled according to probabilities inversely proportional to tensor leverage scores, results evidently become inferior. Meanwhile, we can see that, for the image in Fig. 1, the leveraged sampling favors pixels in the foreground in stead of those in the homogeneous background, generating promising results.

From the definition of the tensor leverage score (to be introduced in Sect. 3.2), we can see that it measures how well a low-dimensional subspace corresponds to coordinate subspaces [43]. Meanwhile, from Theorem 5.3 in [44], we can see that approximating the tensor with the t-SVD framework has less error than with the matrix method. Thus, we can deduce that the low-rank subspace generated by t-SVD is closer to the original tensor than that of the matrix method. Therefore the tensor leverage score captures the importance of the original tensor information better than the matrix counterpart.

The remainder of this paper is organized as follows. Sect. 2 introduces the basic tensor notations and preliminaries for the t-SVD framework. Our main results are given in Sect. 3. We report the experimental results in Sect. 4. Finally, Sect. 5 discusses limitations and draws some conclusions.

2. Notations and preliminaries

Throughout this paper, we use lowercase letters, e.g., x, boldface lowercase letters, e.g., $\mathbf{x}$, boldface upper-case letters, e.g., $\mathbf{X}$, and boldface calligraphic letters, e.g., $\mathcal{X}$, to denote scalars, vectors, matrices, and tensors, respectively. $\mathcal{X}_{ijk}$ is used to denote the (i, j, k)-th element of a given third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. The k-th frontal slice of $\mathcal{X}$ is denoted as $\mathcal{X}^{(k)}$ (or $\mathcal{X}(:, :, k)$). In this paper, we use $\widehat{\mathcal{X}}$ to denote the transformed tensor by performing one-dimensional discrete Fourier transform (DFT) along the mode-3 fibers (tubes) of $\mathcal{X}$.[2]

Definition 1 *(Tensor conjugate transpose [38]).* The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted as $\mathcal{A}^\top \in \mathbb{R}^{n_2 \times n_1 \times n_3}$, is obtained by conjugate transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through n_3:

[2] $\widehat{\mathcal{X}}$ can be efficiently computed by using the fast Fourier transform (FFT) and obtained via the Matlab command `fft`$(\mathcal{X}, [], 3)$.

$$(\mathcal{A}^\top)^{(1)} = (\mathcal{A}^{(1)})^\top,$$
$$(\mathcal{A}^\top)^{(i)} = (\mathcal{A}^{(n_3+2-i)})^\top, i = 2, \cdots, n_3,$$

where $(\cdot)^\top$ denotes the matrix conjugate transpose.

Definition 2 *(T-prod [38]).* The t-prod $\mathcal{C} = \mathcal{A} * \mathcal{B}$ between $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 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 $\mathcal{C}(i, j, :)$ is given by

$$\mathcal{C}(i, j, :) = \sum_{k=1}^{n_2} \mathcal{A}(i, k, :) \circledast \mathcal{B}(k, j, :),$$

where $\circledast$ denotes the circular convolution between two same sized tubes.

Definition 3 *(Special tensor [38]).* The **identity** tensor $\mathcal{I} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ is the tensor whose first frontal slice is the $n_1 \times n_1$ identity matrix, and whose other frontal slices are all zeros. A tensor $\mathcal{Q} \in \mathbb{C}^{n_1 \times n_1 \times n_3}$ is **orthogonal** if it satisfies $\mathcal{Q}^\top * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^\top = \mathcal{I}$. A tensor $\mathcal{A}$ is called **f-diagonal** if each frontal slice $\mathcal{A}^{(i)}$ is a diagonal matrix.

Definition 4 *(Block diagonal form of third-order tensor [5]).* Let $\overline{\mathcal{A}}$ denote the block-diagonal matrix of the tensor $\mathcal{A}$ in the Fourier domain, i.e.,

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

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 $\mathcal{A}$, i.e., $\overline{\mathcal{A}^\top} = \overline{\mathcal{A}}^\top$. We also have that

$$\mathcal{A} * \mathcal{B} = \mathcal{C} \Leftrightarrow \overline{\mathcal{A}}\ \overline{\mathcal{B}} = \overline{\mathcal{C}}.$$

Definition 5 *(Tensor Singular Value Decomposition (t-SVD) [37,38]).* For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the t-SVD of $\mathcal{A}$ is given by

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top,$$

where $*$ is the t-prod, $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor, and entries in $\mathcal{S}$ are called the singular values of $\mathcal{A}$.

Definition 6 *(Tensor tubal-rank [39]).* The tensor tubal-rank 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 the tensor tubal-rank is that it is the largest rank of all the frontal slices of $\hat{\mathcal{A}}$ in Fourier domain. If we say a third-order tensor $\mathcal{A}$ is of full tubal-rank, it means $r = \min\{n_1, n_2\}$.

Definition 7 *(Tensor operator and tensor operator norm [39]).* Tensor operators are denoted by Calligraphic letters. Let $\mathcal{L}$ be a tensor operator mapping a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ to a tensor $\mathcal{B} \in \mathbb{R}^{n_4 \times n_2 \times n_3}$ via the t-product as follows:

$$\mathcal{B} = \mathcal{L}(\mathcal{A}) = \mathcal{L} * \mathcal{A},$$

where $\mathcal{L} \in \mathbb{R}^{n_4 \times n_1 \times n_3}$. Then, the tensor operator norm of $\mathcal{L}$ is defined as

$$\|\mathcal{L}\|_{\text{op}} = \sup_{\{\mathcal{X} | \|\mathcal{X}\|_F \le 1\}} \|\mathcal{L}(\mathcal{X})\|_F.$$

Definition 8 *(Tensor spectral norm and infinity norm [39]).* The tensor spectral norm $\|\mathcal{A}\|$ is defined as the largest singular value of $\mathcal{A}$ and the tensor infinity norm $\|\mathcal{A}\|_\infty$ is defined as the largest absolute value of $\mathcal{A}$.

Moreover, we have that $\|\mathcal{A}\| = \|\overline{\mathcal{A}}\|$.

Definition 9 *(Inner product of tensors [39]).* If $\mathcal{A}$ and $\mathcal{B}$ are third-order tensors of same size $n_1 \times n_2 \times n_3$, then the inner product between $\mathcal{A}$ and $\mathcal{B}$ is defined as the following,

$$\langle \mathcal{A}, \mathcal{B} \rangle = \frac{1}{n_3} \text{trace}(\overline{\mathcal{B}}^\top \overline{\mathcal{A}}),$$

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

Subsequently, the tensor Frobenius norm is defined as

$$\|\mathcal{A}\|_F = \langle \mathcal{A}, \mathcal{A} \rangle^{\frac{1}{2}},$$

and we can easily verify that $\|\mathcal{A}\|_F = \frac{1}{\sqrt{n_3}} \|\overline{\mathcal{A}}\|_F$.

Definition 10 *(Tubal nuclear norm (TNN) [39]).* The tubal nuclear norm of a tensor $\mathcal{A}$, denoted as $\|\mathcal{A}\|_{\text{TNN}}$, is the sum of singular values of all the frontal slices of $\hat{\mathcal{A}}$. In particular,

$$\|\mathcal{A}\|_{\text{TNN}} = \|\overline{\mathcal{A}}\|_*.$$

Definition 11 *(Tensor basis [39]).* The tensor column basis $\vec{e}_i \in \mathbb{R}^{n_1 \times 1 \times n_3}$ is defined with one entry equaling 1 and the rest equaling zero. 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 tensor tube basis is $\mathring{e}_i \in \mathbb{R}^{1 \times 1 \times n3}$ is defined with the i-th entry equaling to 1 and rest equaling to 0.

Definition 12 *(l_{2*} norm of tensor column [39]).* Let $\vec{x}$ be an $n_1 \times 1 \times n_3$ tensor column, we define an l_{2*} norm on it as follows

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

3. Main results

In this section, we consider the low-tubal-rank tensor completion problem via solving the following tensor nuclear norm minimization model. The TNN minimization model is formulated as follows

$$\begin{aligned} &\min_{\mathcal{X}} \ \|\mathcal{X}\|_{\text{TNN}} \\ &\text{s.t. } \mathcal{X}_{ijk} = \mathcal{M}_{ijk}, (i,j,k) \in \Omega, \end{aligned} \tag{3}$$

where Ω is the index set of observed entries. Different from previous works, in which Ω is generated via uniformly random sampling, our work considers the situation where the elements are randomly chosen according to a specific biased distribution. Before giving the specific distribution, the definition of the tensor leverage score is needed.

Definition 13 *(Tensor Leverage Score).* Let the reduced t-SVD of a tensor $\mathcal{M}$ be $\mathcal{U} * \mathcal{S} * \mathcal{V}^\top$. Its (normalized) leverage scores $\mu_i(\mathcal{M})$ for the i-th tensor row $\mathcal{M}(i,:,:)$ and $\nu_j(\mathcal{M})$ for the j-th tensor column $\mathcal{M}(:,j,:)$ are defined as

$$\begin{aligned} \mu_i(\mathcal{M}) &:= \frac{n_1}{r} \left\| \mathcal{U}^\top * \vec{e}_i \right\|_{2*}^2, \ i = 1, 2, \cdots, n_1, \\ \nu_j(\mathcal{M}) &:= \frac{n_2}{r} \left\| \mathcal{V}^\top * \vec{e}_j \right\|_{2*}^2, \ j = 1, 2, \cdots, n_2. \end{aligned} \tag{4}$$

The tensor incoherent condition in (25) of [39] represents the case that all leverage scores are bounded by μ_0.

Theorem 1. *Suppose $\mathcal{M}$ is an $n_1 \times n_2 \times n_3$ tensor and its reduced t-SVD is given by $\mathcal{U} * \mathcal{S} * \mathcal{V}^\top$ 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 $\mathcal{V} \in \mathbb{R}^{n_2 \times r \times n_3}$. There is a universal constant $c_0 > 0$ such that, if each element (i,j,k) is independently observed with probability p_{ijk}, and p_{ijk} satisfies*

$$\begin{aligned} p_{ijk} &\geq \min \left\{ c_0 \frac{(\mu_i(\mathcal{M}) + \nu_j(\mathcal{M}))r \log^2((n_1+n_2)n_3)}{\min\{n_1, n_2\}}, 1 \right\}, \\ p_{ijk} &\geq \frac{8}{\min\{n_1, n_2\}^8 n_3^{11}}, \end{aligned}$$

then $\mathcal{M}$ is the unique minimizer to (3) with probability at least $1 - 5(n_1+n_2)^{-10}$.

Remark 1. The second inequality in Theorem 1 only requires all p_{ijk} being non-zero. The specific value of $\frac{8}{\min\{n_1,n_2\}^8 n_3^{11}}$ comes from the scaling of some inequalities in our proof of Theorem 1. This lower bound could be adjusted while the probability of successful recovery will also change accordingly.

Remark 2. If p_{ijk} in Theorem 1 are all the same, indicating the uniformly random sampling, and all tensor leverage scores are upper bounded by μ_0, our theoretical result in Theorem 1 would reduce to the existing guarantees for incoherent tensors with uniformly random sampling in [39].

Remark 3. If $n_3 = 1$, Theorem 1 conforms to Theorem 2 in [17] and the road map of our analysis also largely follows the line in [17]. However, as mentioned in [39], when $n_3 > 1$, under the leveraged sampling of the tensor, the problem does not reduce to slice-wise matrix completion under leveraged sampling in the Fourier domain since the measurements are coupled. Thus, the main difference brought in by the tensor case makes our analysis more difficult than [17], as the sampling is in the original domain while the tensor-tensor product is computed in the Fourier transformed domain.

We can easily deduce that the minimal expected number of observed elements for exact recovery of $\mathcal{M}$ in (3) is $\sum_{ijk} p_{ijk}$, and we have

$$\begin{aligned}\sum_{ijk} p_{ijk} &\geq \max\left\{ \sum_{i,j,k} \frac{8}{(\min\{n_1,n_2\})^8 n_3^{11}} \ , \ c_0 \frac{r\log^2((n_1+n_2)n_3)}{\min\{n_1,n_2\}} \sum_{i,j,k} (\mu_i(\mathcal{M}) + \nu_j(\mathcal{M})) \right\} \\ &= 2c_0 n_3 r \max\{n_1,n_2\} \log^2\left((n_1+n_2)n_3\right).\end{aligned}$$

3.1. Proof of Theorem 1

The proof of Theorem 1 follows by **(i)** verifying that $\mathcal{M}$ is the unique minimum tubal nuclear norm solution to (3) under specific conditions, and **(ii)** then validating that such conditions are met w.h.p. under the conditions of Theorem 1.

Without loss of generality, we assume $n_1 = n_2 = n$. 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,

$$\mathcal{R}_\Omega(\mathcal{Z}) = \sum_{i,j,k} \frac{1}{p_{ijk}} \delta_{ijk} \mathcal{Z}_{ijk} \vec{e}_i * \mathring{e}_k * \vec{e}_j^\top, \tag{5}$$

where $\delta_{ijk} = \mathbf{1}_{(i,j,k)\in\Omega}$ is a random variable and $\mathbf{1}_{(\cdot)}$ is the indicator function.

Before continuing, let's define two projections $\mathcal{P}_T$ and $\mathcal{P}_{T^\perp}$, given that the t-SVD of $\mathcal{M}$ in (3) is $\mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, as follows,

$$\begin{aligned}\mathcal{P}_T(\mathcal{Z}) &= \mathcal{U} * \mathcal{U}^\top * \mathcal{Z} + \mathcal{Z} * \mathcal{V} * \mathcal{V}^\top - \mathcal{U} * \mathcal{U}^\top * \mathcal{Z} * \mathcal{V} * \mathcal{V}^\top, \\ \mathcal{P}_{T^\perp}(\mathcal{Z}) &= \mathcal{Z} - \mathcal{P}_T(\mathcal{Z}) = (\mathcal{I} - \mathcal{U} * \mathcal{U}^\top) * \mathcal{Z} * (\mathcal{I} - \mathcal{V} * \mathcal{V}^\top),\end{aligned}$$

where $\mathcal{I} \in \mathbb{R}^{n\times n\times n_3}$ is the identity tensor. With projections defined above, the tensor space $\mathbb{R}^{n\times n\times n_3}$ is orthogonally decomposed into T and $T^\perp$ and they satisfy $\mathbb{R}^{n\times n\times n_3} = T \oplus T^\perp$.

Then, we give the following Proposition, to derive conditions required for verifying whether $\mathcal{M}$ is the unique minimum solution to (3).

Proposition 1. *The tensor $\mathcal{M}$ is the minimizer to (3) if the following conditions hold: (i) $\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\|_{op} \leq \frac{1}{2}$; (ii) there exists a tensor $\mathcal{Y}$ such that $\mathcal{P}_\Omega(\mathcal{Y}) = \mathcal{Y}$, $\|\mathcal{P}_T(\mathcal{Y}) - \mathcal{U} * \mathcal{V}^\top\|_F \leq \frac{1}{n^4 n_3^5}$, and $\|\mathcal{P}_{T^\perp}(\mathcal{Y})\| \leq \frac{1}{2}$.*

To prove Proposition 1, we restate some facts borrowed from [39] as follows

$$\begin{cases} \|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\text{TNN}} = n_3 \left\langle \mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{P}_{T^\perp}(\mathcal{Z}) \right\rangle, \\ \|\mathcal{M}\|_{\text{TNN}} = n_3 \left\langle \mathcal{U} * \mathcal{V}^\top + \mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{M} \right\rangle, \\ \|\mathcal{U} * \mathcal{V}^\top + \mathcal{U}_\perp * \mathcal{V}_\perp^\top\| = 1. \end{cases}$$

Meanwhile, the following lemma is needed. Its proof can be found in Appendix.

Lemma 1. *Suppose $\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\|_{op} \leq \frac{1}{2}$. Denote the smallest p_{ijk} as p_{min}. Then for any $\mathcal{Z}$ such that $\mathcal{P}_\Omega(\mathcal{Z} = 0)$, we have*

$$\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{TNN} > \frac{\sqrt{2n_3 p_{min}}}{2} \|\mathcal{P}_T(\mathcal{Z})\|_F.$$

Given any $\mathcal{Z} \in \mathbb{R}^{n\times n\times n_3}$ such that $\mathcal{P}_\Omega(\mathcal{Z}) = 0$ and the t-SVD of $\mathcal{P}_{T^\perp}(\mathcal{Z})$ to be $\mathcal{P}_{T^\perp}(\mathcal{Z}) = \mathcal{U}_\perp * \mathcal{S}_\perp * \mathcal{V}_\perp$, we have

$$\begin{aligned}
\|\mathcal{M}+\mathcal{Z}\|_{\text{TNN}} &\geq n_3 \left\langle \mathcal{U} * \mathcal{V}^\top + \mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{M}+\mathcal{Z} \right\rangle \\
&= \|\mathcal{M}\|_{\text{TNN}} + n_3 \left\langle \mathcal{U} * \mathcal{V}^\top + \mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{Z} \right\rangle \\
&= \|\mathcal{M}\|_{\text{TNN}} + n_3 \left\langle \mathcal{U} * \mathcal{V}^\top, \mathcal{P}_T(\mathcal{Z}) \right\rangle + n_3 \left\langle \mathcal{U}_\perp * \mathcal{V}_\perp^\top, \mathcal{P}_{T^\perp}(\mathcal{Z}) \right\rangle \\
&= \|\mathcal{M}\|_{\text{TNN}} + n_3 \left\langle \mathcal{U} * \mathcal{V}^\top - \mathcal{P}_T(\mathcal{Y}), \mathcal{P}_T(\mathcal{Z}) \right\rangle + n_3 \left\langle \mathcal{U}_\perp * \mathcal{V}_\perp^\top - \mathcal{P}_{T^\perp}(\mathcal{Y}), \mathcal{P}_{T^\perp}(\mathcal{Z}) \right\rangle \\
&= \|\mathcal{M}\|_{\text{TNN}} + \left\langle \overline{\mathcal{U}}\ \overline{\mathcal{V}}^\top - \overline{\mathcal{P}_T(\mathcal{Y})}, \overline{\mathcal{P}_T(\mathcal{Z})} \right\rangle + \|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\text{TNN}} - \langle \mathcal{P}_{T^\perp}(\mathcal{Y}), \mathcal{P}_{T^\perp}(\mathcal{Z}) \rangle \\
&\geq \|\mathcal{M}\|_{\text{TNN}} - \|\overline{\mathcal{U}}\ \overline{\mathcal{V}}^\top - \overline{\mathcal{P}_T(\mathcal{Y})}\|_F \|\overline{\mathcal{P}_T(\mathcal{Z})}\|_F + \|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\text{TNN}} - \|\overline{\mathcal{P}_{T^\perp}(\mathcal{Y})}\| \|\overline{\mathcal{P}_{T^\perp}(\mathcal{Z})}\|_*
\end{aligned}$$

Then, considering that the $\|\overline{\mathcal{P}_{T^\perp}(\mathcal{Z})}\|_*$ is exactly the TNN of $\mathcal{P}_{T^\perp}(\mathcal{Z})$, we have

$$\begin{aligned}
\|\mathcal{M}+\mathcal{Z}\|_{\text{TNN}} &\geq \|\mathcal{M}\|_{\text{TNN}} - \|\overline{\mathcal{U}}\ \overline{\mathcal{V}}^\top - \overline{\mathcal{P}_T(\mathcal{Y})}\|_F \|\overline{\mathcal{P}_T(\mathcal{Z})}\|_F + \|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\text{TNN}} - \|\overline{\mathcal{P}_{T^\perp}(\mathcal{Y})}\| \|\overline{\mathcal{P}_{T^\perp}(\mathcal{Z})}\|_* \\
&= \|\mathcal{M}\|_{\text{TNN}} - n_3 \|\mathcal{U} * \mathcal{V}^\top - \mathcal{P}_T(\mathcal{Y})\|_F \|\mathcal{P}_T(\mathcal{Z})\|_F + \|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\text{TNN}} - \|\mathcal{P}_{T^\perp}(\mathcal{Y})\| \|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\text{TNN}} \\
&\geq \|\mathcal{M}\|_{\text{TNN}} - \frac{1}{n^4 n_3^4} \|\mathcal{P}_T(\mathcal{Z})\|_F + \frac{1}{2} \|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{\text{TNN}} \\
&> \|\mathcal{M}\|_{\text{TNN}}.
\end{aligned}$$

The above inequality is similar to [39] and interested readers can refer to it for detailed derivations. So far, for any $\mathcal{X} \neq \mathcal{M}$ obeying $\mathcal{P}_\Omega(\mathcal{X}-\mathcal{M}) = 0$, we have $\|\mathcal{X}\|_{\text{TNN}} > \|\mathcal{M}\|_{\text{TNN}}$, which proves $\mathcal{M}$ is the unique minimizer of (3).

In the subsequent part, we will validate the first condition in Proposition 1, construct a tensor dual certificate $\mathcal{Y}$, and prove the constructed $\mathcal{Y}$ satisfies both requirements of the second condition in Proposition 1.

Lemma 2. *If* $p_{ijk} \geq \min\{\frac{28\beta \log(nn_3)(\mu_i+\nu_j)r}{3n}, 1\}$ *for all* (i,j,k)*, then*

$$\|\mathcal{P}_T \mathcal{R}_\Omega \mathcal{P}_T - \mathcal{P}_T\|_{op} \leq \frac{1}{2}$$

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

Lemma 2, whose proof can be seen in Appendix, shows that condition 1 in Proposition 1 is satisfied w.h.p. if p_{ijk} satisfies conditions in Theorem 1.

The construction of the dual certificate follows the Golfing Scheme introduced by Gross [45] using the iterative strategy in [14,17,16,39]. Let Ω be a union of smaller sets Ω_t such that $\Omega = \cup_{t=1}^{t_0} \Omega_t$ where $t_0 = \lfloor 5\log(nn_3) \rfloor$. For each t, it is assumed that

$$\mathbb{P}[(i,j,k) \in \Omega_t] = q_{ijk} := 1 - (1-p_{ijk})^{\frac{1}{t_0}},$$

and it is not difficult to verify that it's equivalent to our original Ω. Define $\mathcal{R}_{\Omega_t}(\mathcal{Z})$ as follows

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

Set $\mathcal{W}_0 = \mathbf{0}$ being a zero tensor and for each $t = 1, 2, \cdots, t_0$,

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

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

For $t = 0, 1, \cdots, t_0$, 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

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

Applying Lemma 2 with Ω replaced by Ω_t, with probability at least $1 - 2(nn_3)^{(1-\frac{3c_0}{140})}$, we obtain that

$$\|\mathcal{D}_t\|_F \leq \|\mathcal{P}_T - \mathcal{P}_T \mathcal{R}_{\Omega_T} \mathcal{P}_T\|_{\text{op}} \|\mathcal{D}_{t-1}\|_F \leq \frac{1}{2} \|\mathcal{D}_{t-1}\|_F.$$

Applying the above inequality recursively with $t := t_0, t_0 - 1, \cdots, 1$, gives

$$\|\mathcal{P}_T(\mathcal{Y}) - \mathcal{U} * \mathcal{V}^\top\|_F \leq (\frac{1}{2})^{t_0} \|\mathcal{U} * \mathcal{V}^\top\|_F \leq \frac{\sqrt{r}}{n^5 n_3^5} \leq \frac{1}{n^4 n_3^5}.$$

In order to proceed, we need more definitions.

Definition 14 *(Tensor $\mu(\infty,2*)$ norm).* The $\mu(\infty,2*)$ norm of a tensor $\mathcal{Z}\in\mathbb{R}^{n_1\times n_2\times n_3}$ is defined as

$$\|\mathcal{Z}\|_{\mu(\infty,2*)} := \max\left\{\max_i \sqrt{\frac{n}{\mu_i r}\sum_{b,k}\mathcal{Z}_{ibk}^2}, \max_j \sqrt{\frac{n}{\nu_j r}\sum_{a,k}\mathcal{Z}_{ajk}^2}\right\}$$

Definition 15 *(Tensor $\mu(\infty)$ norm).* The $\mu(\infty)$ norm of a tensor $\mathcal{Z}\in\mathbb{R}^{n_1\times n_2\times n_3}$ is defined as

$$\|\mathcal{Z}\|_{\mu(\infty)} := \max_{i,j}\left|\mathcal{Z}_{ijk}\sqrt{\frac{n}{\mu_i r}}\sqrt{\frac{n}{\nu_j r}}\right|.$$

Before we continue, three technical lemmas, whose proof can be found in Appendix, are shown as follows.

Lemma 3. *If $p_{ijk}\geq\min\{c_0\frac{(\mu_i+\nu_j)r\log^2(2nn_3)}{n},1\}$, and $\mathcal{Z}\in R^{n\times n\times n_3}$, then for any constant $c\geq\frac{4}{\log(2nn_3)}$, we have*

$$\|\mathcal{R}_\Omega(\mathcal{Z})-\mathcal{Z}\|\leq\frac{c}{\sqrt{c_0}}\left(\|\mathcal{Z}\|_{\mu(\infty)}+\|\mathcal{Z}\|_{\mu(\infty,2*)}\right)$$

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

Lemma 4. *If $p_{ijk}\geq\min\{c_0\frac{(\mu_i+\nu_j)r\log^2(2nn_3)}{n},1\}$, for some constant c_1 making $\frac{c_1}{c_0}\leq\frac{\log(2nn_3)}{48}$, and $\mathcal{Z}\in R^{n\times n\times n_3}$, then*

$$\|\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z})-\mathcal{P}_T(\mathcal{Z})\|_{\mu(\infty,2*)}\leq\frac{1}{2}(\|\mathcal{Z}\|_{\mu(\infty)}+\|\mathcal{Z}\|_{\mu(\infty,2*)})$$

holds with probability at least $1-2n^{2-c_1}n_3^{-(c_1-1)}$.

Lemma 5. *If $p_{ijk}\geq\min\{c_0\frac{(\mu_i+\nu_j)r\log^2(2nn_3)}{n},1\}$, and $\mathcal{Z}\in R^{n\times n\times n_3}$, then*

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

holds with probability at least $1-2n^{-(\frac{3c_0\log(2nn_3)}{124}-2)}n_3^{1-\frac{3c_0\log(2nn_3)}{124}}$.

By the construction, $\mathcal{Y}$ can be rewritten as $\sum_{t=1}^{t_0}(\mathcal{R}_{\Omega_T}\mathcal{P}_T)(\mathcal{D}_{t-1})$. It follows that

$$\|\mathcal{P}_{T^\perp}(\mathcal{Y})\|\leq\sum_{t=1}^{t_0}\|(\mathcal{R}_{\Omega_T}-\mathcal{I})\mathcal{P}_T(\mathcal{D}_{t-1})\|.$$

Then, we apply Lemma 3, and have that, w.h.p.,

$$\|\mathcal{P}_{T^\perp}(\mathcal{Y})\|\leq\frac{c}{\sqrt{c_0}}(\sum_{t=1}^{t_0}\|\mathcal{D}_{t-1}\|_{\mu(\infty)}+\sum_{t=1}^{t_0}\|\mathcal{D}_{t-1}\|_{\mu(\infty,2*)}).$$

After applying Lemma 5, we have, w.h.p.,

$$\|\mathcal{D}_{t-1}\|_{\mu(\infty)}=\|(\mathcal{P}_T-\mathcal{P}_T\mathcal{R}_{\Omega_{t-1}}\mathcal{P}_T)\mathcal{D}_{t-2}\|_{\mu(\infty)}\leq(\frac{1}{2})^{t-1}\|\mathcal{U}*\mathcal{V}^\top\|_{\mu(\infty)}.$$

Also, applying Lemma 4, we have, w.h.p.,

$$\begin{aligned}\|\mathcal{D}_{t-1}\|_{\mu(\infty,2*)}&=\|(\mathcal{P}_T-\mathcal{P}_T\mathcal{R}_{\Omega_{t-1}}\mathcal{P}_T)\mathcal{D}_{t-2}\|_{\mu(\infty,2*)}\\&\leq\frac{1}{2}\|\mathcal{D}_{t-2}\|_{\mu(\infty)}+\frac{1}{2}\|\mathcal{D}_{t-2}\|_{\mu(\infty,2*)}\\&\leq(\frac{1}{2})^{t-1}\|\mathcal{U}*\mathcal{V}^\top\|_{\mu(\infty)}+\frac{1}{2}\|\mathcal{D}_{t-2}\|_{\mu(\infty,2*)}\\&\leq(t-1)(\frac{1}{2})^{t-1}\|\mathcal{U}*\mathcal{V}^\top\|_{\mu(\infty)}+(\frac{1}{2})^{t-1}\|\mathcal{U}*\mathcal{V}^\top\|_{\mu(\infty,2)}.\end{aligned}$$

It follows that

Algorithm 1 Two-phase sampling for tensor completion.

Input: Tubal-rank r, m, $\beta \in [0,1]$

Step 1: Acquire the first index set Ω_1 by sampling uniformly without replacement such that $|\Omega_1| = \beta m$. Compute t-SVD of $\mathcal{P}_\Omega(M), \widetilde{U}\widetilde{\Sigma}\widetilde{V}^T$, and estimate the tensor leverage scores $\tilde{\mu}_i$ and $\tilde{v}_j$.

Step 2: Obtain the second index set Ω_2 by sampling without replacement according to the distribution (6). Then, solve the minimization problem in (7).

Output: The completed tensor $\hat{\mathcal{M}}$.

$$\begin{aligned}\|\mathcal{P}_{T^\perp}(\mathcal{Y})\| &\le \frac{c}{\sqrt{c_0}}\Big(\sum_{t=1}^{t_0}(t)(\frac{1}{2})^{t-1}\|\mathcal{U}*\mathcal{V}^\top\|_{\mu(\infty)} + \sum_{t=1}^{t_0}(\frac{1}{2})^{t-1}\|\mathcal{U}*\mathcal{V}^\top\|_{\mu(\infty,2*)}\Big)\\ &\le 4\frac{c}{\sqrt{c_0}}\|\mathcal{U}*\mathcal{V}^\top\|_{\mu(\infty)} + 2\frac{c}{\sqrt{c_0}}\|\mathcal{U}*\mathcal{V}^\top\|_{\mu(\infty,2*)}.\end{aligned}$$

With c_0 large enough, we can conclude that

$$\|\mathcal{P}_{T^\perp}(\mathcal{Y})\| \le 6\frac{c}{\sqrt{c_0}} \le \frac{1}{2}.$$

As we have shown in the above argument, if p_{ijk} satisfies the inequalities in Theorem 1 for all (i,j,k), conditions in Proposition 1 are met with probability at least $1-5(n_1+n_2)^{-10}$. Then, the proof of Theorem 1 follows directly from Proposition 1.

3.2. Two-phase adaptive sampling

We have seen that an arbitrary $n\times n\times n_3$ tensor with tubal-rank r can be exactly restored with $O(nn_3r\log^2(nn_3))$ elements which are sampled according to the tensor leverage scores. A pertinent question is how to conduct the random sampling according to the probability described in Theorem 1 when tensor leverage scores are not known a priori. In this part, we proposed a two-phase adaptive sampling strategy, which is greatly inspired by [17].

Suppose m elements of $\mathcal{M}\in\mathbb{R}^{n_1\times n_2\times n_3}$ in (3) can be randomly observed, we can estimate the tensor leverage scores of the underlying tensor from the set of indices Ω_1 obtained by first sampling uniformly without replacement, where $|\Omega_1| = \beta m$ and $\beta\in[0,1]$ is the proportion of the index number of Ω_1 to the sample budget m. Then, we compute the t-SVD of $\mathcal{P}_{\Omega_1}(\mathcal{M})$, i.e., $\mathcal{P}_{\Omega_1}(\mathcal{M}) = \mathcal{U}_1 * \mathcal{S}_1 * \mathcal{V}_1{}^\top$. Let $\widetilde{\mathcal{U}} = \mathcal{U}_1(:,1:r,:)\in\mathbb{R}^{n\times r\times n_3}$ and $\widetilde{\mathcal{V}} = \mathcal{V}_1(:,1:r,:)\in\mathbb{R}^{n\times r\times n_3}$. Thus, the estimated leverage scores values of $\tilde{\mu}_i$ and values of $\tilde{v}_j$ can be obtained via (4) using $\widetilde{\mathcal{U}}$ and $\widetilde{\mathcal{V}}$. Second, we generate the index set Ω_2, which contains the remaining $(1-\beta)m$ entries of the tensor $\mathcal{M}$, by sampling without replacement with distribution

$$\widetilde{p}_{ijk} \propto \frac{(\widetilde{\mu}_i + \widetilde{v}_j)r\log^2(2\min\{n_1,n_2\}n_3)}{n}. \tag{6}$$

Then, we solve the following problem to obtain the final completion result.

$$\hat{\mathcal{M}} = \arg\min_{\mathcal{X}} \|\mathcal{X}\|_{\mathrm{TNN}} \quad \text{s.t. } \mathcal{P}_{\Omega_1\cup\Omega_2}(\mathcal{X}) = \mathcal{P}_{\Omega_1\cup\Omega_2}(\mathcal{M}). \tag{7}$$

Our two-phase adaptive sampling strategy is summarized in Algorithm 1. Similar to the discussion in [17], if the underlying tensor is incoherent, this algorithm will successfully recover $\mathcal{M}$ if $|\Omega_1| = \Theta(nn_3r\log^2(nn_3))$ with $n = \min\{n_1,n_2\}$. On the contrary, when $\mathcal{M}$ is highly coherent with most of the energy concentrating on only a few elements, the estimation of tensor leverage scores would be seriously affected, resulting in a poor recovery result. Between these two extremes, our two-phase adaptive sampling strategy is believed to generate a better estimation of $\mathcal{M}$ than the entirely uniform random sampling, and the numerical experiments in the subsequent section can support it.

Although the two-phase adaptive sampling strategy is inspired by [17], but we have two key distinctions: i) We use a tensor leverage score, which allows us to handle higher dimensional data better. ii) We propose a specialized two-phase strategy for traffic data by utilizing the periodicity of traffic flow (using the first week or the previous week's data to approximate the leverage score as the first phase, instead of the traditional uniform sampling). This novel design sampling largely enhances the flexibility and practicality of our method for real-world applications.

4. Numerical experiments

In this section, we conduct experiments on synthetic data and real data to verify our theoretical results and the effectiveness of the two-phase sampling strategy.

4.1. Synthetic data

The synthetic tensor $\mathcal{M}$ of size $n\times n\times n_3 = 100\times 100\times 50$ and tubal-rank 2 is constructed via $\mathcal{M} = \mathcal{D}*\mathcal{U}*\mathcal{V}^\top*\mathcal{D}$, where the elements of $\mathcal{U}\in\mathbb{R}^{100\times 2\times 50}, \mathcal{V}\in\mathbb{R}^{2\times 100\times 50}$ are randomly generated obeying the i.i.d. Gaussian $\mathcal{N}(0,1)$ distribution and $\mathcal{D}\in\mathbb{R}^{100\times 100\times 50}$ is a tensor whose first frontal slice is a diagonal matrix with power-law decay, i.e., $\mathcal{D}_{ii1} = i^{-\alpha}$ for $1\le i\le 100$, and whose other frontal slices are zero matrices. We refer to such construction as the *power-law* tensor. The parameter α adjusts the tensor

Table 1
The minimal samples ($\times nn_3 \log(nn_3)$) required for successful recovery of the power-law tensors with different α.

α	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Uniform	1.1	1.3	1.9	1.9	2.9	4	5.3	5.6	7.1	7.3	7.8
Two-phase	1.1	1.1	1.1	1.1	1.3	1.3	1.3	1.4	1.4	1.4	1.5
"Oracle"	1	1	1	1	1	0.9	0.8	0.8	0.7	0.7	0.7

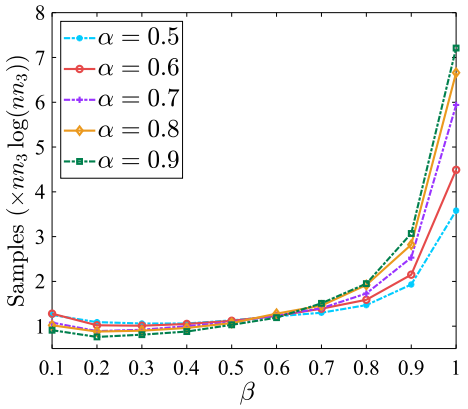

Fig. 2. The minimal samples ($\times nn_3 \log(nn_3)$) required for successful recovery of the power-law tensors with different βs.

leverage scores and the coherence level of $\mathcal{M}$ with $\alpha = 0$ being maximal incoherence $\mu_0 = \Theta(1)$ and $\alpha = 1$ corresponding to maximal coherence $\mu_0 = \Theta(n)$.[3] Finally, we normalize the generated $\mathcal{M}$ to make $\|\mathcal{M}\|_F = 1$.

First, we compare different random sampling schemes, including the uniform sampling, our two-phase sampling strategy, and the leveraged sampling (denoted as "Oracle"), i.e., sampling according to the probability computed in (6) with groundtruth tensor leverage scores. After sampling, the reconstruction is realized via optimizing (3). The minimal samples required for successful recovery, which is defined as when at least 95% of trials have relative errors in the Frobenius norm $\|\mathcal{M} - \hat{\mathcal{M}}\|_F / \|\mathcal{M}\|_F$ not exceeding 0.01, of the simulated tensor with respect to different αs are shown in Table 1. As α becomes larger, the coherence level increases, and most energy of the tensor data would concentrate on a few entries, which are favored by the leveraged sampling. Thus, as α becomes larger, the number of required samples for the successful recovery becomes less through leveraged sampling. Meanwhile, we can see that, as the parameter α varies from 0 to 1, our two-phase sampling strategy achieves comparable results as the leveraged sampling.

Second, we test our two-phase sampling strategy with respect to different values of β. Fig. 2 illustrates the minimal samples required for successful recovery with different values of β. It can be found that the selection of β is robust for the successful recovery in the interval of $[0.1, 0.6]$.

4.2. Color images completion

In this part, we conduct experiments on 20 color images which are of the size $481 \times 321 \times 3$ (width $\times$ height $\times$ color channel) selected from the Berkeley Segmentation Dataset [46]. Compared methods are i) LRMC with uniform sampling [11], denoted as "LRMC (Uniform)", ii) LRMC with a two-phase adaptive sampling strategy [17], denoted as "LRMC (Adaptive)", iii) HaLRTC with random sampling [2], denoted as "HaLRTC (Uniform)", iv) TNN with random sampling [39], denoted as "TNN (Uniform)", v) discrete cosine transformation (DCT) based TNN with random sampling [29], denoted as "DCT-TNN (Uniform)", vi) TNN minimization with leveraged sampling, denoted as "TNN ("Oracle")", and vii) TNN minimization with our two-phase adaptive sampling strategy, denoted as "TNN ("Adaptive")". The peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [47] are computed for quantitative evaluation of the completion results. Higher values of PSNR and SSIM indicate better reconstruction results. In our experiment, 50% of the entries are sampled in total. When conducting the uniformly random sampling and the adaptively leveraged sampling, those observed entries are sampled at once. For a fair comparison, when doing the adaptive two-phase sampling, we set $\beta = 0.6$ and elements sampled in the first phase are randomly extracted from those selected by TNN (Uniform).

We exhibit PSNR values and SSIM values of results in Fig. 3. For better visualization, the vertical axis starts from 20 for the PSNR and 0.75 for the SSIM, respectively. From Fig. 3, we can see an overall trend that i) results by tensor-based methods are better than those by matrix-based methods, and ii) adaptive sampling could help to generate better reconstructions than uniform sampling with the same amount of observed entries. Meanwhile, the performance of our two-phase adaptive sampling is close to "Oracle" leveraged sampling, especially in terms of the SSIM values.

[3] Here, μ_0 is the upper bound of tensor leverage scores and it is indeed the μ_0 in (25) of [39].

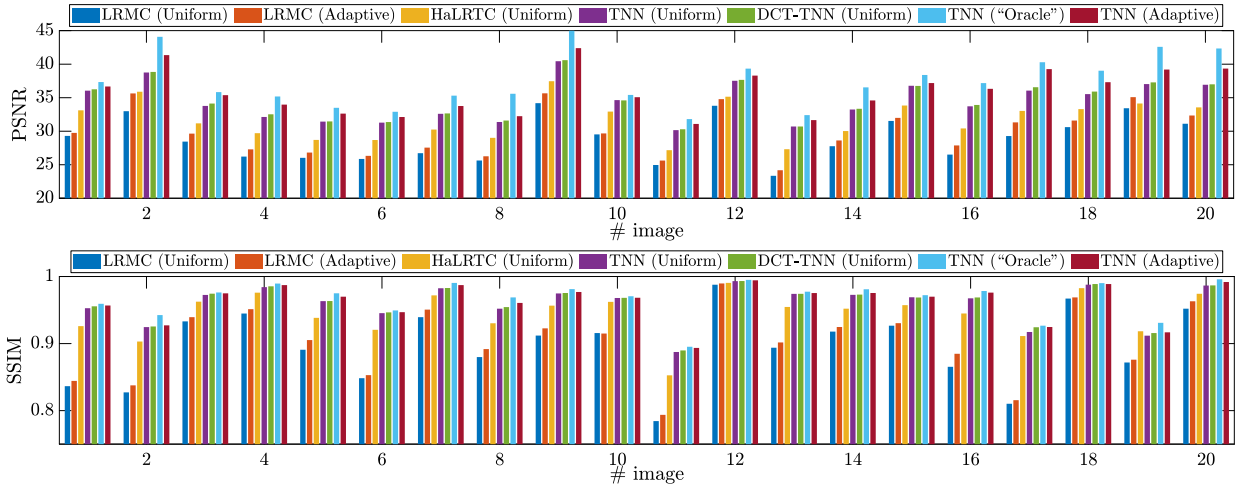

Fig. 3. The PSNR values of the first 20 images and the SSIM values of the last 20 images.

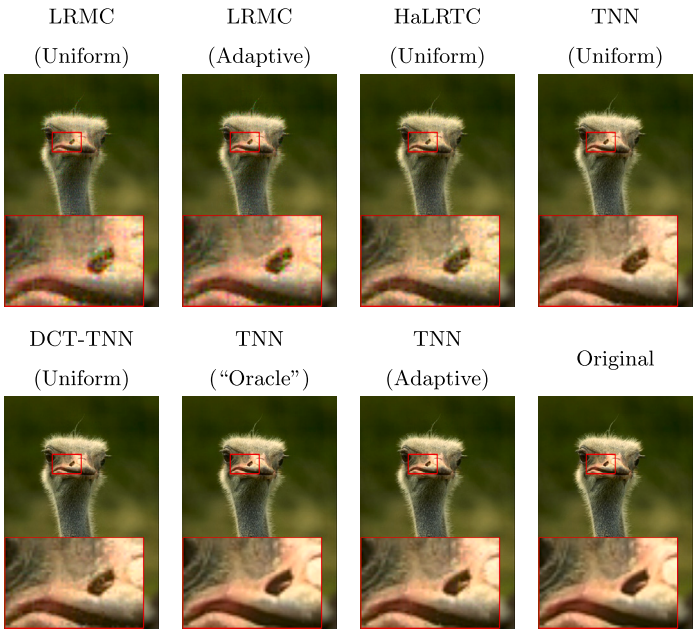

Fig. 4. The completion results of one color image.

The reason our method performs better is that the usage of tensor leverage score allows us to take more information-rich points, as is shown in Fig. 1, which allows for better detail recovery. For example, Fig. 4 displays the reconstruction results on one image. From the enlarged area, the superiority of our methods are obvious as the nostril are restored better. However, for less information-rich points, even if fewer points are selected, it does not have a significant effect on the recovery result. Therefore, with the same total sampling rate, we have better overall results, and our approach is particularly well suited for some downstream tasks that require greater detail. It is worth noting that our adaptive method performs only a little worse than the oracle method, which means that we are able to recover the image fairly well without knowing the leverage score in advance.

4.3. Multispectral image completion

In this section, we conduct experiments on the CAVE data set [48] provided by the Computer Science Department at Columbia University. The data set consists of multispectral images which are of the size $512 \times 512 \times 31$ (spatial height $\times$ spatial width $\times$ spectrum). The third dimension comprising full spectral resolution reflectance data that ranges from 400 nm to 700 nm in 10 nm increments, totaling 31 bands. We selected the mean values of PSNR and SSIM of frontal slices to quantitatively measure the quality of the results. The total sampling rate for all methods is set as 50%, where the sampling points in the first phase of the adaptive

Table 2
The average PSNR and SSIM values of different methods in MSI data.

Method	PSNR	SSIM
Observed	21.71	0.1382
LRMC Uniform	46.32	0.9169
LRMC Adaptive	47.09	0.8935
HaLRTC Uniform	47.66	0.8655
TNN Uniform	55.01	0.9440
DCT-TNN Uniform	54.85	0.9417
TNN Oracle	56.81	0.9421
TNN Adaptive	55.40	0.9436
TNN Adaptive (interpolation)	56.08	0.9452

Table 3
The averaged MAPE and RMSE values of different methods in traffic data.

Methods	MAPE	RMSE
Observe	50.00%	0.6390
LRMC Uniform	2.13%	0.0324
LRMC Adaptive	2.46%	0.0364
HaLRTC Uniform	4.43%	0.0490
TNN Uniform	2.18%	0.0304
DCT-TNN Uniform	2.18%	0.0304
TNN Oracle	1.24%	0.0197
TNN Adaptive-Latest	1.46%	0.0229
TNN Adaptive-First	1.45%	0.0236

sampling accounted for 60% of the total number of sampling points. The TNN Adaptive (interpolated version) method interpolates all the unsampled ones after sampling uniformly in the first stage, and then calculates the leverage score to select the samples for the second stage.

Table 2 provides quantitative measurements of results obtained with different methods. When rotating the images to size $n_3 \times n_1 \times n_2$, the two matrix-based methods (LRMC Uniform and LRMC Adaptive) as well as HaLRTC Uniform performed the worst in terms of both PSNR and SSIM indices. In contrast, tensor methods based on leverage score achieved the best results overall. In particular, the Oracle method excelled in PSNR, while TNN Adaptive (interpolation) achieved the highest SSIM. However, the performance differences among these top methods are slight.

4.4. Traffic data

In this part, we conduct experiments on traffic data [49] (A set of 207 days, 1440 time windows and 21 detection points of traffic speed data) provided by Grenoble traffic Lab (GTL), which can constitute a third-order tensor of size $1440 \times 207 \times 21$. As in practice the leverage score of the week is not known, we designed some methods trying to use the leverage score of historical data. We manually clipped the continuous dataset, selecting 1440 time windows and 21 detection points from the 8th day (inclusive) to the 203rd day (inclusive), and obtained some continuous data set of size $1440 \times 7 \times 21$ as the actual complete test dataset. The Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are computed to quantitatively measure the quality of the results. The sampling rate is 50% for all methods. In this case, as the traffic data exhibit typical cyclicity, we can estimate the tensor leverage scores from previous periods. Thus, we add two alternatives of our method: i) TNN Adaptive-Latest, which means the tensor leverage scores are estimated from last week, and ii) TNN Adaptive-First, which obtains the tensor leverage scores from the first week (not involved in testing). Table 3 gives the quantitative measures of the results obtained by the different methods. We can see that the oracle case of our method achieves the best performance and two alternatives are comparable to the oracle case.

We present MAPE values and RMSE values of results in Fig. 5. From Fig. 5, we observe an overall trend that i) results by leverage score sampling based methods are better than those using uniform sampling, and ii) the sampling method based on this week's leverage score helped reconstruction the most, and iii) the matrix-based method performs worse than other tensor-based methods except HaLRTC Uniform. Meanwhile, the sampling method based on the first and previous week's leverage score also have a significant increase in MAPE and RMSE values, and the difference in the improvement was not significant. This is in large part due to the fact that our tensor leverage score sampling enables sampling to be more focused on data points that differ from their surrounding points. Fig. 6 displays the reconstruction results on 5-th lateral slice of traffic data in week 18. As can be seen from the figure, the three leverage score based methods are much better than the others in some details. Interestingly, the highlighted area that contrasts with the surrounding region is the traffic data of the area around the Centre Commercial Grand'Place during afternoon rush hour, and our advantage in recovering data like that allows for more accurate downstream tasks like traffic jams analysis.

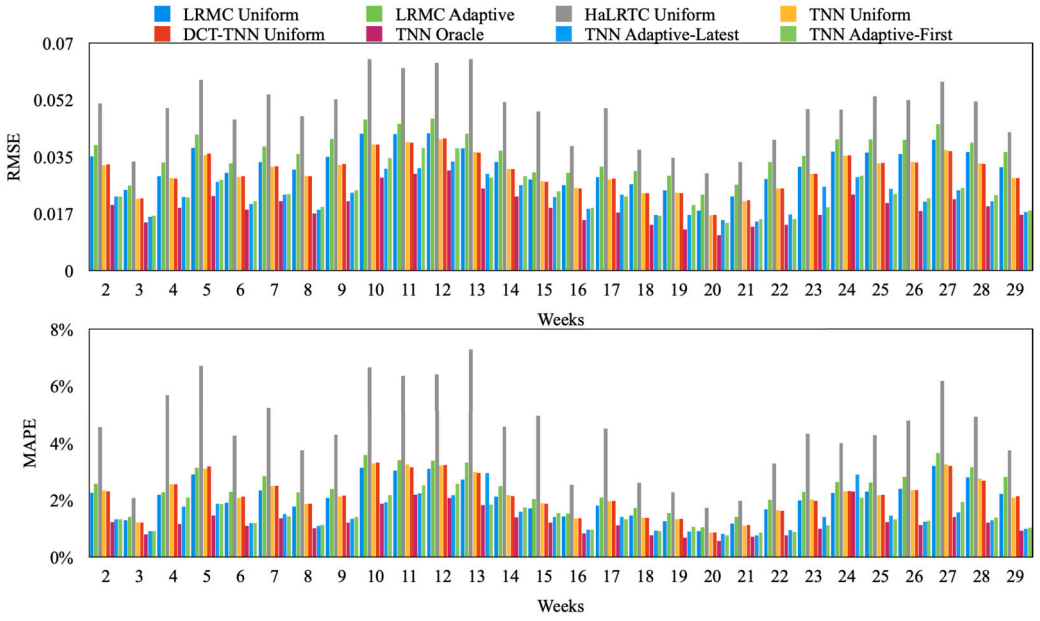

Fig. 5. The MAPE and RMSE values of every week, respectively.

In this experiment, we have given two-phase strategy a very novel meaning. In the first phase, not only can we estimate leverage scores using uniform sampling like what we did in the first three experiments, but we can also utilize historical data to estimate the leverage score of current data in periodic datasets. This insight allows our approach to be used in more realistic scenarios.

5. Concluding remarks

In this work, we define the tensor leverage scores, which can be viewed as the local version of the tensor incoherent condition. Then, we prove that one can exactly recover any low-tubal-rank tensor from as few as $O(rnn_3 \log^2(nn_3))$ random observations via solving a convex optimization problem with high probability. The required condition is that the random selection of entries is made adaptive to a specific distribution and the probability of entries being sampled should be at least a constant times the sum of the tensor leverage scores. A two-phase adaptive sampling strategy follows with our theoretical results. Experiments verify our theory and show that our two-phase adaptive sampling works well.

CRediT authorship contribution statement

Xuan Chen: Writing – original draft, Methodology, Investigation. **Tai-Xiang Jiang:** Writing – original draft, Methodology. **Yexun Hu:** Software, Investigation, Data curation. **Jinjin Yu:** Validation, Software, Investigation. **Michael K. Ng:** Writing – review & editing, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Tai-Xiang Jiang's work was supported in part by Natural Science Foundation of Xinjiang Uygur Autonomous Region (2024D01A18) and the Guanghua Talent Project.

Michael K. Ng's research supported in part by the HKRGC GRF 17201020 and 17300021, and CRF C7004-21GF, and Joint NSFC and RGC NHKU769/21.

Appendix A. Noncommutative Bernstein inequality

First, we introduce the Noncommutative Bernstein Inequality (NBI) [15], which will be frequently utilized throughout this Appendix.

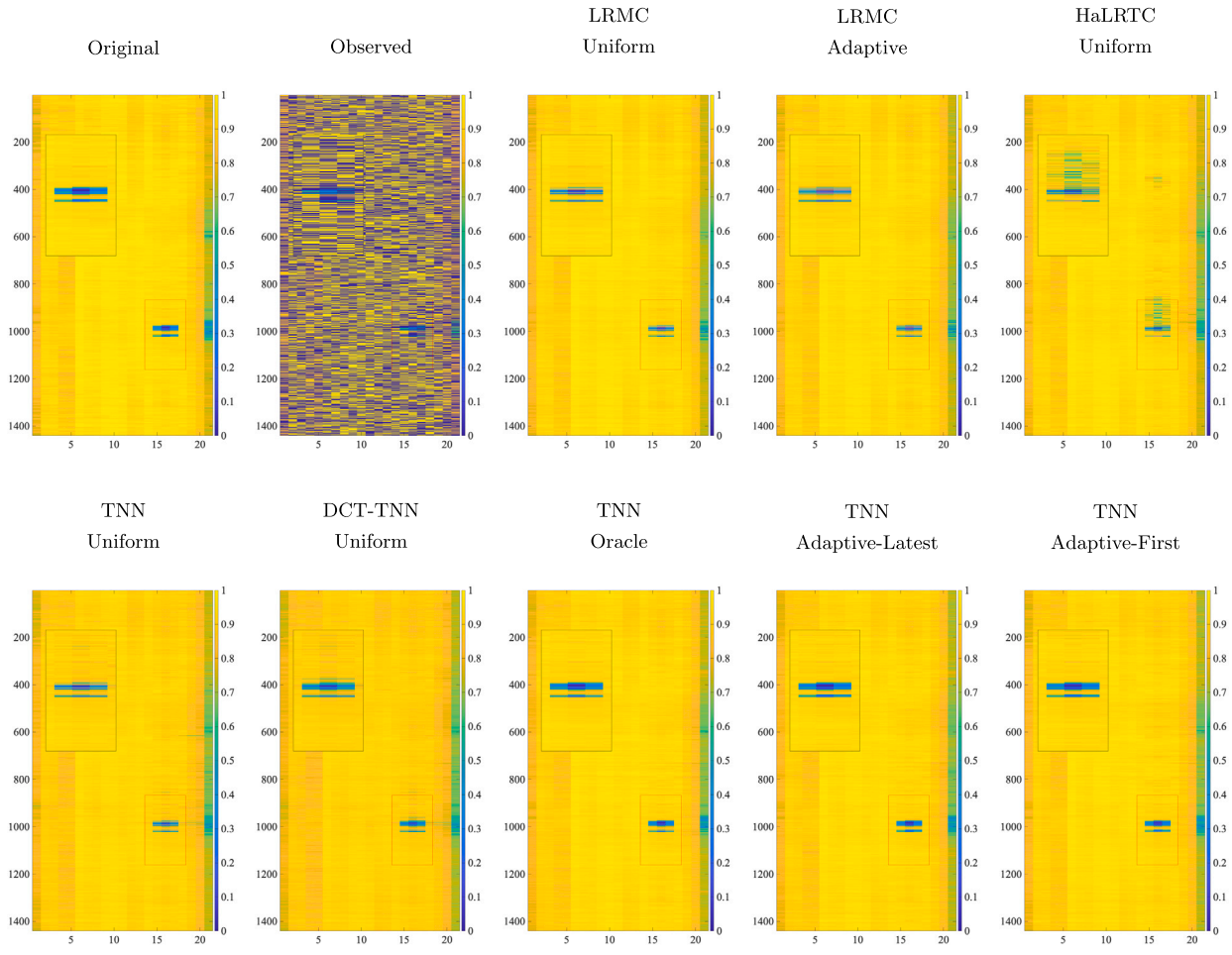

Fig. 6. The 5-th lateral slice of the reconstructions by different methods on the traffic data (Week 18).

Theorem 2 *(Noncommutative Bernstein Inequality [15]). Let* $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_L$ *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

$$\|\mathbf{X}_K\| \le M$$

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

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

If

$$\max\{\|\sum_{k=1}^{L} \mathbf{X}_K \mathbf{X}_K^\top\|, \|\sum_{k=1}^{L} \mathbf{X}_K^\top \mathbf{X}_K\|\} \le \sigma^2$$

and let

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

for any $c > 0$*, then*

$$\mathbb{P}[\|\sum_{k=1}^{L} \mathbf{X}_k\| \ge \tau] \le (d_1 + d_2)^{-(c-1)}.$$

Appendix B. Proof of technical lemmas

In this part, we give the detailed proof of all the lemmas in our main manuscript. For the readers' convenience, we would first restate those lemmas and then prove them.

B.1. Proof of Lemma 1

Lemma 1. *Suppose* $\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\|_{op} \le \frac{1}{2}$. *Denote the smallest* p_{ijk} *as* p_{min}. *Then for any* $\mathcal{Z}$ *such that* $\mathcal{P}_\Omega(\mathcal{Z}=0)$, *we have*

$$\|\mathcal{P}_{T^\perp}(\mathcal{Z})\|_{TNN} > \frac{\sqrt{2n_3p_{min}}}{2}\|\mathcal{P}_T(\mathcal{Z})\|_F.$$

Proof of Lemma 1. Define the operator $\mathcal{R}_\Omega^{\frac{1}{2}} : \mathbb{R}^{n\times n\times n_3} \mapsto \mathbb{R}^{n\times n\times n_3}$ by

$$\mathcal{R}_\Omega^{\frac{1}{2}}(\mathcal{Z}) = \sum_{i,j,k}\frac{1}{\sqrt{p_{ijk}}}\delta_{ijk}\mathcal{Z}_{ijk}\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top.$$

Then

$$\begin{aligned}\|\mathcal{R}_\Omega^{\frac{1}{2}}\mathcal{P}_\mathcal{T}(\mathcal{Z})\|_F &= \sqrt{\langle\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z}),\mathcal{P}_T(\mathcal{Z})\rangle}\\ &= \sqrt{\langle(\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T-\mathcal{P}_T)(\mathcal{Z}),\mathcal{P}_T(\mathcal{Z})\rangle + \langle\mathcal{P}_T(\mathcal{Z}),\mathcal{P}_T(\mathcal{Z})\rangle}\\ &\ge \sqrt{\|\mathcal{P}_T(\mathcal{Z})\|_F^2 - \|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T-\mathcal{P}_T\|_{\mathrm{op}}\|\mathcal{P}_T(\mathcal{Z})\|_F^2}\\ &\ge \frac{1}{\sqrt{2}}\|\mathcal{P}_T(\mathcal{Z})\|_F\end{aligned}$$

Thus

$$\|\mathcal{R}_\Omega^{\frac{1}{2}}\mathcal{P}_\mathcal{T}(\mathcal{Z})\|_F = \|-\mathcal{R}_\Omega^{\frac{1}{2}}\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{Z})\|_F \le \frac{\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{Z})\|_F}{\sqrt{p_{min}}}.$$

Combining the equations gives

$$\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{Z})\|_{TNN} \ge \sqrt{n_3}\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{Z})\|_F \ge \frac{\sqrt{n_3p_{min}}}{\sqrt{2}}\|\mathcal{P}_T(\mathcal{Z})\|_F. \quad \square$$

B.2. Proof of Lemma 2

Lemma 2. *If* $p_{ijk} \ge \min\{\frac{28\beta\log(nn_3)(\mu_i+\nu_j)r}{3n}, 1\}$ *for all* (i,j,k), *then*

$$\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\|_{op} \le \frac{1}{2}$$

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

Before proving Lemma 2, we need a fact borrowed from [39]:

$$\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^{\,\top})\|_F^2 \le \frac{(\mu_i+\nu_j)r}{n}.$$

Proof of Lemma 2. Define the operator $\mathcal{J}_{ijk}$ which maps $\mathcal{Z}$ to $\frac{1}{p_{ijk}}\delta_{ijk}\left\langle\mathcal{Z},\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\right\rangle\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)$. When $p_{ijk} \ge 28\beta\frac{(\mu_i+\nu_j)r\log n}{3n}$, we have

$$\|\mathcal{J}_{ijk} - \frac{1}{n^2n_3}\mathcal{P}_T\|_{\mathrm{op}} = \|\overline{\mathcal{J}_{ijk}} - \frac{1}{n^2n_3}\overline{\mathcal{P}_T}\|_{\mathrm{op}} \le \max\left\{\frac{1}{p_{ijk}}\|\mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^\top)\|_F^2, \frac{1}{n^2n_3}\right\} \le \frac{(\mu_i+\nu_j)r}{p_{ijk}n}$$

On the other hand, observe that we have $\mathbb{E}[\mathcal{J}_{ijk}] = \frac{1}{n^2n_3}\mathcal{P}_T$. Thus, we have

$$\left\|\mathbb{E}\left[(\overline{\mathcal{J}}_{ijk} - \frac{1}{n^2n_3}\overline{\mathcal{P}_T})^2\right]\right\| < \left(\frac{1}{p_{ijk}}\frac{(\mu_i+\nu_j)r}{n}\frac{1}{n^2n_3}\right)\|\mathcal{P}_T\|_{\mathrm{op}} \le \frac{(\mu_i+\nu_j)r}{p_{ijk}n^3n_3}$$

Now, let

$$\tau = \sqrt{\frac{7\beta \log(nn_3) r(\mu_i + \nu_j)}{3p_{ijk}n}} < \frac{1}{2}$$

with some constant $\beta > 1$. Then, we have

$$P\left[\left\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\right\|_{\text{op}} > \tau\right] \le 2nn_3 \exp\left(\frac{-\frac{7}{6}\frac{\beta\log(nn_3)(\mu_i+\nu_j)r}{p_{ijk}n}}{\frac{(\mu_i+\nu_j)r}{p_{ijk}n} + \frac{(\mu_i+\nu_j)r}{6p_{ijk}n}}\right) \le (2nn_3)^{1-\beta}.$$

Therefore, we get

$$P\left[\left\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\right\|_{\text{op}} \le \frac{1}{2}\right] \ge P\left[\left\|\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T - \mathcal{P}_T\right\|_{\text{op}} \le \tau\right] \ge 1 - (2nn_3)^{1-\beta},$$

which finishes the proof. □

B.2.1. Proof of Lemma 3

Lemma 3. *If* $p_{ijk} \ge \min\{c_0\frac{(\mu_i+\nu_j)r\log^2(2nn_3)}{n}, 1\}$, *and* $\mathcal{Z} \in R^{n\times n\times n_3}$, *then for any constant* $c \ge \frac{4}{\log(2nn_3)}$, *we have*

$$\|\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{Z}\| \le \frac{c}{\sqrt{c_0}}\left(\|\mathcal{Z}\|_{\mu(\infty)} + \|\mathcal{Z}\|_{\mu(\infty,2*)}\right)$$

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

Proof of Lemma 3. Let

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

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

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

Notice that $\mathbb{E}[\overline{\mathcal{J}_{ijk}}] = 0$ and $\|\overline{\mathcal{J}_{ijk}}\| \le \frac{1}{p_{ijk}}\|\mathcal{Z}\|_\infty$. Moreover, we have

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

Similarly, $\left\|\mathbb{E}\left[\sum_{i,j,k}\overline{\mathcal{J}_{ijk}}\overline{\mathcal{J}_{ijk}}^{\top}\right]\right\|$ is also bounded in the same way. Then, applying the matrix NBI [15], we have

$$\begin{aligned}\|\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{Z}\| &= \|\overline{\mathcal{R}_\Omega(\mathcal{Z})} - \overline{\mathcal{Z}}\| = \|\sum_{i,j,k}\overline{\mathcal{J}_{(ijk)}}\|_{\text{op}} \\ &\le \sqrt{4c\log(2nn_3)\max\left\{\max_j\sum_{i,k}\frac{1-p_{ijk}}{p_{ijk}}\mathcal{Z}_{ijk}^2, \max_i\sum_{j,k}\frac{1-p_{ijk}}{p_{ijk}}\mathcal{Z}_{ijk}^2\right\}} + c\max_{i,j,k}\frac{|Z_{ijk}|}{p_{ijk}}\log(2nn_3)\end{aligned}$$

holds with probability at least $1 - (2nn_3)^{-(c-1)}$. If $p_{ijk} \ge \min\left\{c_0\frac{(\mu_i+\nu_j)r}{n}\log^2(2nn_3), 1\right\}$ for all (i,j,k) and a sufficiently large constant $c \ge \frac{4}{\log(2nn_3)}$, we further have

$$\|\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{Z}\| \le \frac{c}{\sqrt{c_0}}(\|\mathcal{Z}\|_{\mu(\infty)} + \|\mathcal{Z}\|_{\mu(\infty,2^*)}) \quad \square$$

B.2.2. Proof of Lemma 4

Lemma 4. *If* $p_{ijk} \ge \min\{c_0\frac{(\mu_i+\nu_j)r\log^2(2nn_3)}{n}, 1\}$, *for some constant* c_1 *making* $\frac{c_1}{c_0} \le \frac{\log(2nn_3)}{48}$, *and* $\mathcal{Z} \in R^{n\times n\times n_3}$, *then*

$$\|\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z}) - \mathcal{P}_T(\mathcal{Z})\|_{\mu(\infty,2*)} \le \frac{1}{2}(\|\mathcal{Z}\|_{\mu(\infty)} + \|\mathcal{Z}\|_{\mu(\infty,2*)})$$

holds with probability at least $1 - 2n^{2-c_1}n_3^{-(c_1-1)}$.

Proof of Lemma 4. Consider any b-th tensor column of $\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z})-\mathcal{P}_T(\mathcal{Z})$:

$$\begin{aligned}\sqrt{\frac{n}{\nu_b r}}(\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z})-\mathcal{P}_T(\mathcal{Z}))*\vec{e}_b &= \sqrt{\frac{n}{\nu_b r}}\sum_{i,j,k}(\frac{1}{p_{ijk}}\delta_{ijk}-1)\mathcal{Z}_{ijk}\mathcal{P}_T(\vec{e}_i*\mathring{e}_k*\vec{e}_j^{\top})*\vec{e}_b\\ &= \sum_{i,j,k}\vec{a}_{ijk},\end{aligned}$$

where $\vec{a}_{ijk}\in\mathbb{R}^{n\times 1\times n_3}$ are zero-mean independent tensor columns. Let $\tilde{a}_{ijk}$ be the vectorized column vector of $\vec{a}_{ijk}$. Then for some constant $c_0>0$, we have

$$\|\tilde{a}_{ijk}\| = \|\vec{a}_{ijk}\|_{2*} \le \sqrt{\frac{n}{\nu_b r}}\frac{1-p_{ijk}}{p_{ijk}}\mathcal{Z}_{ijk}\|\mathcal{P}_T(\vec{e}_i*\mathring{e}_k*\vec{e}_j^{\top})*\vec{e}_b\|_{2^*} \le \sqrt{\frac{n}{\nu_b r}}\frac{|\mathcal{Z}_{ijk}|}{p_{ijk}}\|\mathcal{P}_T(\vec{e}_i*\mathring{e}_k*\vec{e}_j^{\top})*\vec{e}_b\|_{2^*}.$$

When $j=b$:

$$\|\tilde{a}_{ijk}\|_2 \le \sqrt{\frac{n}{\nu_b r}}\frac{|\mathcal{Z}_{ijk}|}{p_{ijk}}(\sqrt{\frac{\mu_i r}{n}}+\sqrt{\frac{\nu_b r}{n}}) \le \frac{2}{c_0\log^2(2nn_3)}\|\mathcal{Z}\|_{\mu(\infty)}.$$

When $j\neq b$:

$$\|\tilde{a}_{ijk}\|_2 \le \sqrt{\frac{n}{\nu_b r}}\frac{|\mathcal{Z}_{ijk}|}{p_{ijk}}\sqrt{\frac{\nu_j r}{n}}\sqrt{\frac{\nu_b r}{n}} \le \frac{2}{c_0\log^2(2nn_3)}\|\mathcal{Z}\|_{\mu(\infty)},$$

given $p_{ijk}\ge\min\{\frac{c_0(\mu_i+\nu_j)r\log^2(2nn_3)}{n},1\}$. We also have

$$\begin{aligned}\left\|\mathbb{E}\left[\sum_{i,j,k}\tilde{a}_{ijk}^{\top}\tilde{a}_{ijk}\right]\right\| &= \mathbb{E}\left[\sum_{i,j,k}\|\vec{a}_{ijk}\|_{2*}^2\right]\\ &= \frac{n}{\nu_b r}\frac{1-p_{ijk}}{p_{ijk}}\sum_{i,j,k}\mathcal{Z}_{ijk}^2\|\mathcal{P}_T(\vec{e}_i*\mathring{e}_k*\vec{e}_j^{\top})*\vec{e}_b\|_{2*}^2\\ &= (\sum_{i,j=b,k}+\sum_{i,j\neq b,k})\frac{n}{\nu_b r}\frac{1-p_{ijk}}{p_{ijk}}\mathcal{Z}_{ijk}^2\|\mathcal{P}_T(\vec{e}_i*\mathring{e}_k*\vec{e}_j^{\top})*\vec{e}_b\|_{2*}^2\\ &\le \left(\sum_{i,k}2\frac{\mu_i r+\nu_b r}{n}+\sum_{j\neq b,k}\|\vec{e}_j^{\top}*\mathcal{V}*\mathcal{V}^{\top}*\vec{e}_b\|^2\right)\frac{1-p_{ibk}}{p_{ibk}}\mathcal{Z}_{ibk}^2\frac{n}{\nu_b r}\\ &\le \frac{3}{c_0\log^2(2nn_3)}\|\mathcal{Z}\|_{\mu(\infty,2^*)}^2.\end{aligned}$$

Applying the matrix NBI [15], if $\frac{c_1}{c_0}\le\frac{\log(2nn_3)}{48}$, we have, with probability at least $1-(nn_3)^{-(c_1-1)}$,

$$\begin{aligned}\left\|\sqrt{\frac{n}{\nu_b r}}\left(\mathcal{P}_T\mathcal{R}_\Omega(\mathcal{Z})-\mathcal{P}_T(\mathcal{Z})\right)\vec{e}_b\right\|_{2*} &= \left\|\sum_{i,j,k}\vec{a}_{ijk}\right\|_{2*} = \left\|\sum_{i,j,k}\tilde{a}_{ijk}\right\|_2\\ &\le \sqrt{\frac{12c_1}{c_0\log(2nn_3)}}\|\mathcal{Z}\|_{\mu(\infty,2^*)}+\frac{2c_1}{c_0\log(2nn_3)}\|\mathcal{Z}\|_{\mu(\infty,2)}\\ &\le \frac{1}{2}(\|\mathcal{Z}\|_{\mu(\infty)}+\|\mathcal{Z}\|_{\mu(\infty,2^*)}).\end{aligned}$$

We can also do the same to the tensor rows and get the same bound. Then using a union bound over all the tensor columns and tensor rows, the results hold with probability at least $1-2n^{2-c_1}n_3^{-(c_1-1)}$. □

B.2.3. Proof of Lemma 5

Lemma 5. *If* $p_{ijk}\ge\min\{c_0\frac{(\mu_i+\nu_j)r\log^2(2nn_3)}{n},1\}$, *and* $\mathcal{Z}\in R^{n\times n\times n_3}$, *then*

$$\|(\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T-\mathcal{P}_T)\mathcal{Z}\|_{\mu(\infty)}\le\frac{1}{2}\|\mathcal{Z}\|_{\mu(\infty)}$$

holds with probability at least $1-2n^{-(\frac{3c_0\log(2nn_3)}{124}-2)}n_3^{1-\frac{3c_0\log(2nn_3)}{124}}$.

Proof of Lemma 5. Observe that

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

Let $w_{abc} = \sqrt{\frac{\mu_a r}{n}\frac{\nu_b r}{n}}$. Then, we can write

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

We can compute the following bound:

$$\left|\left\langle \mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^{\top}), \vec{e}_a * \mathring{e}_c * \vec{e}_b^{\top}\right\rangle\right| \le \begin{cases}\frac{\mu_a r}{n} + \frac{\nu_b r}{n}, & \text{if } i = a, j = b,\\ \|\vec{e}_b^{\top} * \mathcal{V} * \mathcal{V}^{\top} * \vec{e}_j\|_{2*}, & \text{if } i = a, j \ne b,\\ \|\vec{e}_i^{\top} * \mathcal{U} * \mathcal{U}^{\top} * \vec{e}_a\|_{2*}, & \text{if } i \ne a, j = b,\\ \|\vec{e}_b^{\top} * \mathcal{V} * \mathcal{V}^{\top} * \vec{e}_j\|_{2*}\|\vec{e}_i^{\top} * \mathcal{U} * \mathcal{U}^{\top} * \vec{e}_a\|_{2*}, & \text{if } i \ne a, j \ne b.\end{cases}$$

If $p_{ijk} \ge \frac{c_0 r(\mu_i+\nu_j)\log^2(2nn_3)}{n}$, it is easy to observe that

$$\left|\mathcal{H}_{ijk,abc}\right| \le \frac{|\mathcal{Z}_{ijk}|}{w_{abc}p_{ijk}}\left|\left\langle \mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^{\top}), \vec{e}_a * \mathring{e}_c * \vec{e}_b^{\top}\right\rangle\right| \le \|\mathcal{Z}\|_{\mu(\infty)}/c_0\log^2(2nn_3).$$

On the other hand, note that

$$\begin{aligned}\left|\mathbb{E}\left[\sum_{i,j,k}\mathcal{H}_{ijk,abc}^2\right]\right| &= \sum_{i,j,k}\mathbb{E}\left[\left(\frac{\delta_{ijk}}{p_{ijk}} - 1\right)^2\right]\frac{\mathcal{Z}_{ijk}^2}{w_{abc}^2}\left\langle \mathcal{P}_T(\vec{e}_i * \mathring{e}_k * \vec{e}_j^{\top}), \vec{e}_a * \mathring{e}_c * \vec{e}_b^{\top}\right\rangle^2\\ &= \sum_{i=a,j=b,k} + \sum_{i=a,j\ne b,k} + \sum_{i\ne a,j=b,k} + \sum_{i\ne a,j\ne b,k}\\ &\le \frac{2\|\mathcal{Z}\|_{\mu(\infty)}^2}{c_0\log^2 2nn_3} + \frac{2\|\mathcal{Z}\|_{\mu(\infty)}^2}{c_0\log^2 2nn_3} + \frac{\|\mathcal{Z}\|_{\mu(\infty)}^2}{c_0\log^2 2nn_3}\\ &= \frac{5\|\mathcal{Z}\|_{\mu(\infty)}^2}{c_0\log^2 2nn_3}.\end{aligned}$$

Then, using the NBI [15], we have

$$\begin{aligned}P\left[\left(\mathcal{P}_T\mathcal{R}_\Omega\mathcal{P}_T(\mathcal{Z}) - \mathcal{P}_T(\mathcal{Z})\right)_{abc} \ge \frac{1}{2}\|\mathcal{Z}\|_{\mu(\infty)}\right] &\le 2\exp\left(\frac{-\frac{1}{8}\|\mathcal{Z}\|_{\mu(\infty)}^2}{\frac{5\|\mathcal{Z}\|_{\mu(\infty)}^2}{c_0\log^2 2nn_3} + \frac{\|\mathcal{Z}\|_{\mu(\infty)}^2}{6c_0\log^2 2nn_3}}\right)\\ &\le 2\exp\left(-\frac{3c_0\log^2 2nn_3}{124}\right)\\ &\le 2(nn_3)^{-\frac{3c_0\log(2nn_3)}{124}}.\end{aligned}$$

Then using the union bound on every (a, b, c)-th entry, we have the inequality holds with probability at least $1 - 2n^{-\left(\frac{3c_0\log(2nn_3)}{124}-2\right)} \times n_3^{1-\frac{3c_0\log(2nn_3)}{124}}$. □

Data availability

Data will be made available on request.

References

[1] N. Komodakis, Image Completion Using Global Optimization, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, IEEE, 2006, pp. 442–452.

[2] J. Liu, P. Musialski, P. Wonka, J. Ye, Tensor completion for estimating missing values in visual data, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 208–220.

[3] X. Li, Y. Ye, X. Xu, Low-rank tensor completion with total variation for visual data inpainting, Proc. AAAI Conf. Artif. Intell. 31 (2017).
[4] X.-L. Zhao, W.-H. Xu, T.-X. Jiang, Y. Wang, M.K. Ng, Deep plug-and-play prior for low-rank tensor completion, Neurocomputing 400 (2020) 137–149.
[5] Z. Zhang, G. Ely, S. Aeron, N. Hao, M. Kilmer, Novel methods for multilinear data completion and de-noising based on tensor-svd, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 3842–3849.
[6] L. Zhuang, J.M. Bioucas-Dias, Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (3) (2018) 730–742.
[7] W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, H. Zhang, L. Zhang, Non-local meets global: an integrated paradigm for hyperspectral image restoration, IEEE Trans. Pattern Anal. Mach. Intell. (2020).
[8] Z. Lin, M. Chen, Y. Ma, The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices, arXiv preprint arXiv:1009.5055, 2010.
[9] S. Boyd, N. Parikh, E. Chu, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Now Publishers Inc., 2011.
[10] J.-F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. 20 (4) (2010) 1956–1982.
[11] E.J. Candès, B. Recht, Exact matrix completion via convex optimization, Found. Comput. Math. 9 (6) (2009) 717–772.
[12] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Rev. 52 (3) (2010) 471–501.
[13] E.J. Candès, T. Tao, The power of convex relaxation: near-optimal matrix completion, IEEE Trans. Inf. Theory 56 (5) (2010) 2053–2080.
[14] B. Recht, A simpler approach to matrix completion, J. Mach. Learn. Res. 12 (12) (2011).
[15] D. Gross, Y.-K. Liu, S.T. Flammia, S. Becker, J. Eisert, Quantum state tomography via compressed sensing, Phys. Rev. Lett. 105 (15) (2010) 150401.
[16] Y. Chen, Incoherence-optimal matrix completion, IEEE Trans. Inf. Theory 61 (5) (2015) 2909–2923.
[17] Y. Chen, S. Bhojanapalli, S. Sanghavi, R. Ward, Completing any low-rank matrix, provably, J. Mach. Learn. Res. 16 (1) (2015) 2999–3034.
[18] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Rev. 51 (3) (2009) 455–500.
[19] Z. Long, Y. Liu, L. Chen, C. Zhu, Low rank tensor completion for multiway visual data, Signal Process. 155 (2019) 301–316.
[20] Q. Song, H. Ge, J. Caverlee, X. Hu, Tensor completion algorithms in big data analytics, ACM Trans. Knowl. Discov. Data 13 (1) (2019) 6.
[21] H.A. Kiers, Towards a standardized notation and terminology in multiway analysis, J. Chemom. J. Chemom. Soc. 14 (3) (2000) 105–122.
[22] Q. Zhao, L. Zhang, A. Cichocki, Bayesian CP factorization of incomplete tensors with automatic rank determination, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1751–1763.
[23] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, S.-I. Amari, Bayesian robust tensor factorization for incomplete multiway data, IEEE Trans. Neural Netw. Learn. Syst. 27 (4) (2016) 736–748.
[24] L.R. Tucker, Some mathematical notes on three-mode factor analysis, Psychometrika 31 (3) (1966) 279–311.
[25] I.V. Oseledets, Tensor-train decomposition, SIAM J. Sci. Comput. 33 (5) (2011) 2295–2317.
[26] J.A. Bengua, H.N. Phien, H.D. Tuan, M.N. Do, Efficient tensor completion for color image and video recovery: low-rank tensor train, IEEE Trans. Image Process. 26 (5) (2017) 2466–2479.
[27] Q. Zhao, G. Zhou, S. Xie, L. Zhang, A. Cichocki, Tensor ring decomposition, arXiv preprint arXiv:1606.05535, 2016.
[28] W. Wang, V. Aggarwal, S. Aeron, Efficient low rank tensor ring completion, in: IEEE International Conference on Computer Vision (ICCV), 2017, pp. 5697–5705.
[29] C. Lu, X. Peng, Y. Wei, Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 5996–6004.
[30] S. Gandy, B. Recht, I. Yamada, Tensor completion and low-n-rank tensor recovery via convex optimization, Inverse Probl. 27 (2) (2011) 025010.
[31] B. Huang, C. Mu, D. Goldfarb, J. Wright, Provable low-rank tensor recovery, Optim. Online 4252 (2) (2014) 455–500.
[32] C. Mu, B. Huang, J. Wright, D. Goldfarb, Square deal: lower bounds and improved relaxations for tensor recovery, in: International Conference on Machine Learning (ICML), PMLR, 2014, pp. 73–81.
[33] J.-F. Cai, J. Li, D. Xia, Provable tensor-train format tensor completion by Riemannian optimization, arXiv preprint arXiv:2108.12163, 2021.
[34] H. Huang, Y. Liu, J. Liu, C. Zhu, Provable tensor ring completion, Signal Process. 171 (2020) 107486.
[35] K. Braman, Third-order tensors as linear operators on a space of matrices, Linear Algebra Appl. 433 (7) (2010) 1241–1253.
[36] M.E. Kilmer, C.D. Martin, Factorization strategies for third-order tensors, Linear Algebra Appl. 435 (3) (2011) 641–658.
[37] N. Hao, M.E. Kilmer, K. Braman, R.C. Hoover, Facial recognition using tensor-tensor decompositions, SIAM J. Imaging Sci. 6 (1) (2013) 437–463.
[38] M.E. Kilmer, K. Braman, N. Hao, R.C. Hoover, Third-order tensors as operators on matrices: a theoretical and computational framework with applications in imaging, SIAM J. Matrix Anal. Appl. 34 (1) (2013) 148–172.
[39] Z. Zhang, S. Aeron, Exact tensor completion using t-svd, IEEE Trans. Signal Process. 65 (6) (2017) 1511–1526.
[40] G. Song, M.K. Ng, X. Zhang, Robust tensor completion using transformed tensor singular value decomposition, Numer. Linear Algebra Appl. 27 (3) (2020) e2299.
[41] X.-Y. Liu, S. Aeron, V. Aggarwal, X. Wang, M.-Y. Wu, Adaptive sampling of RF fingerprints for fine-grained indoor localization, IEEE Trans. Mob. Comput. 15 (10) (2015) 2411–2423.
[42] A. Wang, Z. Jin, X. Li, Coherent low-tubal-rank tensor completion, in: 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), 2017, pp. 518–523.
[43] R. Murray, J. Demmel, M.W. Mahoney, N.B. Erichson, M. Melnichenko, O.A. Malik, L. Grigori, P. Luszczek, M. Derezinski, M.E. Lopes, T. Liang, H. Luo, J. Dongarra, Randomized numerical linear algebra: a perspective on the field with an eye to software, Tech. Rep. UCB/EECS-2023-19, Feb 2023, http://www2.eecs.berkeley.edu/Pubs/TechRpts/2023/EECS-2023-19.html.
[44] M.E. Kilmer, L. Horesh, H. Avron, E. Newman, Tensor-tensor algebra for optimal representation and compression of multiway data, Proc. Natl. Acad. Sci. 118 (2021), https://api.semanticscholar.org/CorpusID:235768323.
[45] D. Gross, Recovering low-rank matrices from few coefficients in any basis, IEEE Trans. Inf. Theory 57 (3) (2011) 1548–1566.
[46] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proc. 8th Int'l Conf. Computer Vision, vol. 2, 2001, pp. 416–423.
[47] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[48] F. Yasuma, T. Mitsunaga, D. Iso, S. Nayar, Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum, Tech. Rep., Nov 2008.
[49] C.C. de Wit, F. Morbidi, L.L. Ojeda, A.Y. Kibangou, I. Bellicot, P. Bellemain, Grenoble traffic lab: an experimental platform for advanced traffic monitoring and forecasting [applications of control], IEEE Control Syst. Mag. 35 (3) (2015) 23–39, https://doi.org/10.1109/MCS.2015.2406657.