Exact Tensor Completion Beyond Isotropy and Invertibility

Li Ge
Shanghai Jiao Tong University
geli2000an@sjtu.edu.cn

Lin Chen
Stevens Institute of Technology
lchen53@stevens.edu

Yudong Chen
University of Wisconsin-Madison
yudongchen@cs.wisc.edu

Xue Jiang*
Shanghai Jiao Tong University
xuejiang@sjtu.edu.cn

Abstract

In this work, a tensor completion problem is studied, which aims to perfectly recover the tensor from partial observations. The existing theoretical guarantee requires the involved transform to be orthogonal, which hinders its applications. In this paper, jumping out of the constraints of isotropy and invertibility for the first time, the theoretical guarantee of exact tensor completion with arbitrary linear transforms is established by directly operating the tensors in the transform domain. With the enriched choices of transforms, we theoretically disclose why slim transforms outperform their square counterparts, providing support for existing works on experimental excellence of slim transforms. Our model and analysis greatly enhance the flexibility of tensor completion, supported by extensive experimental results.

1 INTRODUCTION

Tensors, or multidimensional arrays, are data structures extensively applied in computer vision (Panagakis et al., 2021), signal processing (Sidiropoulos et al., 2017) and machine learning (Stephan et al., 2017). With the volume ever-growing, a substantial proportion of data inherently possesses a multiway nature, e.g., RGB/RGB-D/multispectral data (Zhao et al., 2024), light field images (He et al., 2024), medical images (Yin, 2018), network traffic flow (Li et al., 2023), etc., which leads us to organize them into tensors and enables efficient compression, transmission and recovery(Kolda and Bader, 2009).

Commonly, it is not easy to satisfactorily compress or recover a tensor without any prior information. However, a specific property or prior of tensors that sheds light on tensor processing is the low-rank property. Enjoying the analogy with matrices (De Lathauwer and De Moor, 1998), it is conceivable that low-rank tensors will have great advantages being represented, compressed, or recovered.

Unfortunately, unlike the matrix case, in which the rank-related theory has been well established, low-rank tensors can be unexpectedly difficult to identify since the concept "rank" is not well-defined for tensors. Over the past decades, several definitions of tensor ranks have been proposed, each with its limitations. For instance, the well-known CP rank (Kolda and Bader, 2009) encounters numerical problems and has no tractable convex relaxation, i.e., nuclear norm, which is a useful tool powered by convex optimization and has proven its success in matrix recovery(Chen et al., 2013), matrix completion (Candès and Tao, 2010) and robust principle component analysis (RPCA) (Candès et al., 2011). Another category of definitions stems from the tensor multilinear rank, in which the rank of tensor $\mathcal{X}$ is a vector with each entry being the rank of a specified unfolding of $\mathcal{X}$. Tensor decompositions possessing such a multilinear rank include Tucker (Liu et al., 2013), tensor train (Oseledets, 2011), tensor ring (Zhao et al., 2016), and the fully-connected tensor network (Zheng et al., 2021). However, these decompositions are all sub-optimal (Mu et al., 2014) and their completion guarantees are worse

* Corresponding author. This work was supported by the National Natural Science Foundation of China under Grant 62571315. (Corresponding author: Xue Jiang.)

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

than the matrix completion approach. [1]

In recent years, leveraging the tensor-tensor product and tensor singular value decomposition (t-SVD), a new tensor rank, i.e., tensor tubal rank is developed for 3-order tensors (Zhang and Aeron, 2017) and can be extended to high order cases (Qin et al., 2022). By viewing tensor fibers of the third dimension as tubes, a crucial concept in the methods using t-SVD is the applied transform. A good transform should strengthen the low-rankness of the tensors, i.e., making the tensor more favorable for recovering in the transform domain than in the original domain. First, in (Zhang and Aeron, 2017), the tensor nuclear norm (TNN) is induced by discrete Fourier transform (DFT). Equipped with TNN, tensor completion and tensor RPCA (Lu et al., 2020) enjoy the exact recovery guarantee. Further, by extending DFT to any invertible orthogonal transform $\mathbf{L}$ and encapsulating $\mathbf{L}(\cdot)$ into the induced tensor-tensor product $\star_{\mathbf{L}}$, the exact recovery can still be guaranteed with enriched choices of the transforms (Lu et al., 2019; Lu, 2021).

However, increasing experimental phenomena from recent works (Jiang et al., 2020; Wang et al., 2021; Luo et al., 2022; Jiang et al., 2023) indicate that compared to the square orthogonal transforms, "slim" and non-invertible transforms that map the original tensor into bigger transform domains can achieve better recovery accuracy and flexibility (some of the transforms are self-adaptive) in tensor-related tasks. Unfortunately, the existing theoretical result requires isotropy and invertibility (both are properties of orthogonal/unitary transforms), thus fails to guarantee the recovery under these non-invertible transforms. Moreover, the determination of the transforms is empirical and lacks theoretical justification as to why slim transforms are better. This work remedies both of these problems.

In this paper, we studies the problem of tensor completion, a fundamental task that plays a crucial role in various applications such as data impainting, reducing transmission bandwidth, and reliable data compression (Song et al., 2019). Corresponding to the aforementioned existing works, our main contributions are as follows:

- Jumping out of the constraints of isotropy and invertibility, we prove that exact tensor completion can be achieved with **arbitrary** linear transforms (Theorem 3.4) at a near-optimal sampling rate, which provides theoretical support for these empirical works. To that aim, different from existing proofs (Zhang and Aeron, 2017; Lu et al., 2019), we develop a set of tensor operators that directly work on the transform domain, removing the redundant domain transition (Section 2).
- Among the enriched choices of the transforms that enjoy the exact completion guarantee, our results theoretically show why slim transforms outperform square transforms by showing a sampling rate reduction (Theorem 3.7, and full analysis in Appendix E), supporting the existing experimental phenomena (Jiang et al., 2020; Wang et al., 2021; Luo et al., 2022; Jiang et al., 2023).
- We specifically explored the cascaded transform family, showing that such simple yet effective designs can achieve comparable performances to the widely-used FTNN method through extensive experiments on random data and visual data.

Our proof of the main theorem is completely new and challenging, as the operators are mingled with the transform $\mathbf{T}$ (see Section 3.4 for detailed explanation). Since we require neither isotropy nor invertibility upon the linear transform, our model and proof are applicable to much more general transforms. Furthermore, the ingredients in our proof can be readily applied to other tensor related tasks, such as tensor recovery from linear transforms and tensor RPCA. Substantial experiments on random data and visual data validate the superiority of the proposed method.

2 DEFINE TENSOR OPERATORS ON THE TRANSFORM DOMAIN

In this section, we first introduce notations used throughout the paper. Then we show the difficulties encountered using existing definitions (Zhang and Aeron, 2017; Lu et al., 2019), which inspire us to work directly in the transform domain. [2]

Notations Lowercase letters (e.g. a) denote scalars; their boldface versions (e.g. $\mathbf{a}$) denote vectors. Matrices and tensors are denoted by capital boldface letters (e.g. $\mathbf{A}$) and calligraphic letters (e.g. $\mathcal{A}$), respectively. $(\cdot)^{\mathrm{T}}, (\cdot)^{\mathrm{H}}$ and $(\cdot)^{\dagger}$ denote transpose, adjoint/Hermitian transpose and pseudo-inverse, respectively. For indexing vectors, matrices, and tensors, we use $\mathcal{A}_{ijk}$ or $\mathcal{A}(i,j,k)$ to denote the (i,j,k)th entry of $\mathcal{A}$ and use : to represent a full selection in the corresponding mode. $\langle\cdot,\cdot\rangle$ denotes the inner product. $\|\cdot\|_{\mathrm{F}}, \|\cdot\|_{\infty}$ stand for the Frobenius and infinity norm of vectors/matrices/tensors, respectively. $\|\cdot\|/\|\cdot\|_{\mathrm{op}}$ stands

[1] To better see this, consider that in building exact completion bounds for these tensor decompositions, all matrix unfoldings must be exactly completed. However, using the matrix completion framework, completing only one of these unfoldings will suffice.

[2] Throughout this paper, we consider 3-order tensor, as t-SVD type methods are proposed for 3-order tensors. However, our model and proof can be readily extended to handle higher order tensors, using the extension "high-order t-SVD" proposed in (Qin et al., 2022).

for spectral/operator norm (induced by $\|\cdot\|_{\mathrm{F}}$ in both domains unless specified). $\|\cdot\|_*$ is the nuclear norm. $\|\cdot\|_{1\mapsto 2}, \|\cdot\|_{2\mapsto\infty}$ are the largest ℓ_2 norm of the matrix/tensor columns and rows, respectively, and $\|\cdot\|_{\infty,2} = \max(\|\cdot\|_{1\mapsto 2}, \|\cdot\|_{2\mapsto\infty})$. $\mathrm{Ber}(p)$ denotes Bernoulli sampling with rate p. By *with high probability* (w.h.p.), we mean with probability at least $1-c_1(n_1+n_2)^{-c_2}$, where c_1, c_2 are universal constants.

For operators, $\times_3$ denotes the mode-3 product. $\texttt{vec}(\cdot), \texttt{bdiag}(\cdot)$ and $\texttt{fold}(\cdot)$ denote vectorization, organizing the frontal slices of a tensor into a block diagonal matrix and its inverse operation, respectively. $\sigma_{\max}(\cdot)/\sigma_{\min}(\cdot)$ stands for taking the maximum/minimum singular value of a matrix. $(\mathcal{A})_+$ denotes taking the maximum of 0 and each entry of $\mathcal{A}$. $x \vee y$ means taking the larger of x, y.

2.1 Difficulties Encountered with Existing Definitions

Assume an arbitrary linear transform $T(\cdot)$ is given by $\mathbf{T} \in \mathbb{C}^{N_3\times n_3}$ and $T(\mathcal{A}) = \mathcal{A} \times_3 \mathbf{T}$ for $\mathcal{A} \in \mathbb{C}^{n_1\times n_2\times n_3}$, let us first attempt to use the transform-induced tensor-tensor product $\star_{\mathbf{T}}$ to define tensor nuclear norm following the methodology in (Lu et al., 2019). With transform-induced $\star_{\mathbf{T}}$ and subsequent t-SVD. One can define the tensor spectral norm as $\|\mathcal{A}\| := \|\texttt{bdiag}(T(\mathcal{A}))\|$ for a tensor $\mathcal{A}$. It is tempting to define the tensor nuclear norm as the dual of the tensor spectral norm, still following (Lu et al., 2019):

$$\|\mathcal{A}\|_* := \sup_{\|\mathcal{B}\|\le 1} \langle \mathcal{A}, \mathcal{B}\rangle = \sup_{\|\texttt{bdiag}(T(\mathcal{B}))\|\le 1} \langle \mathcal{A}, \mathcal{B}\rangle.$$

Now the problem of losing the norm-preserving property occurs since $\forall \mathcal{A}, \mathcal{B}, \langle \mathcal{A}, \mathcal{B}\rangle = \langle T(\mathcal{A}), T(\mathcal{B})\rangle$ holds only when $\mathbf{T}$ is orthogonal/unitary (or with some constant l). Therefore, although such a manner of definition that wraps the transform into the element-element operation does bring convenience and elegance in form, the ability for generalization is hindered. In this paper, we show that the operations to push tensors from the orginal domain to the transform domain and the pull-back can be redundant, i.e., tensors can be directly operated in the transform domain. To this aim, tensor operators without the transform involved are defined in the remainder of this section.

2.2 Tensor Operators Defined Directly on the Transform Domain

First, we introduce the tensor-tensor product defined directly on the transform domain.

Definition 2.1 (**Tensor-tensor product**). [3] For two tensors $\mathcal{A} \in \mathbb{C}^{n_1\times n_2\times n_3}, \mathcal{B} \in \mathbb{C}^{n_2\times l\times n_3}$, tensor-tensor product between $\mathcal{A}$ and $\mathcal{B}$ is denoted as $\mathcal{C} = \mathcal{A} \star \mathcal{B} \in \mathbb{C}^{n_1\times l\times n_3}$, such that $\mathcal{C}(:,:,k) = \mathcal{A}(:,:,k)\mathcal{B}(:,:,k), k = 1,\dots,n_3$.

It is evident that such tensor-tensor product is simply the matrix-matrix product on each frontal slice thus a few relative definitions can be given as expected.

Definition 2.2 (**Tensor (conjugate) transpose**). Given a tensor $\mathcal{A} \in \mathbb{C}^{n_1\times n_2\times n_3}$, the tensor transpose $\mathcal{A}^{\mathrm{T}} \in \mathbb{C}^{n_2\times n_1\times n_3}$ can be defined following $\mathcal{A}^{\mathrm{T}}(:,:,k) = [\mathcal{A}(:,:,k)]^{\mathrm{T}}$. The tensor conjugate transpose $\mathcal{A}^{\mathrm{H}}$ can be defined as $\mathcal{A}^{\mathrm{H}}(:,:,k) = [\mathcal{A}(:,:,k)]^{\mathrm{H}}$.

Definition 2.3 (**Identity tensor**). The identity tensor $\mathcal{I} \in \mathbb{C}^{n_1\times n_1\times n_3}$ is defined such that each frontal slice of $\mathcal{I}$ is an identity matrix $\mathbf{I} \in \mathbb{C}^{n_1\times n_1}$.

Definition 2.4 (**Unitary tensor**). A tensor $\mathcal{U} \in \mathbb{C}^{n_1\times n_1\times n_3}$ is called unitary if $\mathcal{U}\star\mathcal{U}^{\mathrm{H}} = \mathcal{I}$.

Definition 2.5 (**T-SVD**). Tensor SVD (t-SVD) of $\mathcal{A} \in \mathbb{C}^{n_1\times n_2\times n_3}$is performing matrix SVD for each frontal slice of $\mathcal{A}$. The process of computing t-SVD is given in Algorithm 1.

Algorithm 1 T-SVD

Input: $\mathcal{A} \in \mathbb{C}^{n_1\times n_2\times n_3}$
Output: $\mathcal{U} \in \mathbb{C}^{n_1\times n_1\times n_3}, \mathcal{S} \in \mathbb{C}^{n_1\times n_2\times n_3}, \mathcal{V} \in \mathbb{C}^{n_2\times n_2\times n_3}$
for $k = 1$ **to** n_3 **do**
  $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \mathrm{SVD}(\mathcal{A}(:,:,k))$;
  $\mathcal{U}(:,:,k) = \mathbf{U}, \mathcal{S}(:,:,k) = \mathbf{S}, \mathcal{V}(:,:,k) = \mathbf{V}$;
end for

After performing t-SVD defined as above, one can verify that $\mathcal{U}, \mathcal{V}$ decomposed are both unitary according to Definition 2.4.

With t-SVD, we are ready to further define the tensor rank and the tensor spectral norm.

Definition 2.6 (**Tensor tubal rank**). For $\mathcal{A} \in \mathbb{C}^{n_1\times n_2\times n_3}$ with t-SVD $\mathcal{A} = \mathcal{U}\star\mathcal{S}\star\mathcal{V}^{\mathrm{H}}$, the tensor tubal rank of $\mathcal{A}$ is defined as the number of nonzero tubes of $\mathcal{S}$, i.e., $\mathrm{rank}(\mathcal{A}) = \#\{i, \mathcal{S}(i,i,:) \neq \mathbf{0}\}$.

Sharing the analogy with matrices, we also have skinny t-SVD as $\mathcal{A} = \mathcal{U}_r \star \mathcal{S}_r \star \mathcal{V}_r^{\mathrm{H}}$ with $\mathcal{U}_r \in \mathbb{C}^{n_1\times r\times n_3}, \mathcal{S}_r \in \mathbb{C}^{r\times r\times n_3}, \mathcal{V}_r \in \mathbb{C}^{n_2\times r\times n_3}$, where $r = \mathrm{rank}(\mathcal{A})$. Since the skinny t-SVD explicitly unveils the subspace of $\mathcal{A}$, we use it when referring to "t-SVD" throughout the paper.

Definition 2.7 (**Tensor spectral norm**). For a tensor $\mathcal{A} \in \mathbb{C}^{n_1\times n_2\times n_3}$ with t-SVD $\mathcal{A} = \mathcal{U}\star\mathcal{S}\star\mathcal{V}^{\mathrm{H}}$, the tensor spectral norm is the maximum value of $\mathcal{S}$, i.e., $\|\mathcal{A}\| = \max_{i,k} \mathcal{S}(i,i,k)$.

[3] Such slice-wise operation has also appeared in (Kernfeld et al., 2015). However, the purposes differ greatly: (Kernfeld et al., 2015) defines "face-wise product" for the convenience of tensor product whereas it still operates the tensors in the original domain (please see its Definition 4.2 and Algorithm 2, 3). While in this work, tensors are directly operated in the transform domain.

It can be verified that the tensor spectral norm of $\mathcal{A}$ in Definition 2.7 is the operator norm induced by Frobenius norm in both domains as we treat "$\mathcal{A} \star (\cdot)$" as an operator and use the fact that the matrix operator norm is the matrix spectral norm.

Now we can define the tensor nuclear norm as the dual norm of the tensor spectral norm:[4]

Definition 2.8 (**Tensor nuclear norm**)**.** For a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ with t-SVD $\mathcal{A} = \mathcal{U} \star \mathcal{S} \star \mathcal{V}^{\mathrm{H}}$, the tensor nuclear norm is defined as $\|\mathcal{A}\|_* = \sum_{k=1}^{n_3} \sum_{i=1}^{r} \mathcal{S}(i,i,k)$.

3 EXACT COMPLETION GUARANTEE WITH ARBITRARY LINEAR TRANSFORMS

Equipped with tools directly built on the transform domain, we now consider completing a tensor with an arbitrary linear transform. In this section, under certain tensor incoherence conditions, a theoretical guarantee to achieve exact tensor completion with arbitrary linear transforms is established. Furthermore, we show why slim transforms outperform their square counterparts, providing theoretical justification for recent emerging experimental works.

3.1 The Optimization Program

The program we consider can be expressed as

$$\min_{\mathcal{X}} \|T(\mathcal{X})\|_* \quad \text{s.t. } \mathcal{S}_\Omega(\mathcal{X}) = \mathcal{S}_\Omega(\mathcal{M}), \tag{1}$$

where $T : \mathbb{C}^{n_3} \mapsto \mathbb{C}^{N_3}$ is a linear injection, $T(\mathcal{X}) = \mathcal{X} \times_3 \mathbf{T}, \mathbf{T} \in \mathbb{C}^{N_3 \times n_3}$. $\mathcal{X}, \mathcal{M} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, $\mathcal{S}_\Omega$ is a sampling operator with $\Omega \sim \mathrm{Ber}(p)$. $[\mathcal{S}_\Omega(\mathcal{X})]_{ijk} = \mathcal{X}_{ijk}, (i,j,k) \in \Omega$ and equaling 0 elsewhere. When $\mathbf{T}$ is the discrete Fourier transform (DFT), Program 1 regresses to the optimization problem considered in existing TNN works such as(Zhang and Aeron, 2017).

Program 1 can be solved by the standard ADMM framework, with details provided in Appendix F.

Program 1 is motivated by a natural idea: we wish to recover tensor $\mathcal{X}$ from partial observations Ω (often containing a small fraction of entries), while the intrinsic low-rank property of $\mathcal{X}$ itself might not be significant enough. Thus, we seek to minimize its nuclear norm in the transform domain characterized by a transform $\mathbf{T}$. Obviously, the transform $\mathbf{T}$ plays a direct and crucial role in the success of the completion. Initially, this is only considered in the context of DFT representing the idea that convolution can well capture the structure along the third mode of the visual data (Zhang and Aeron, 2017). The transform is further generalized to any invertible orthogonal transforms, e.g. discrete cosine transform (DCT) and discrete sine transform (DST), while still maintaining the exact recovery guarantee (Lu et al., 2019).

However, the requirements on invertibility and orthogonality limit our choices of transforms. It has been shown that the low-rank property might become more pronounced when the transform maps $\mathcal{X}$ into a larger space, which is not invertible ($N_3 > n_3$) (Jiang et al., 2020; Luo et al., 2022; Jiang et al., 2023). Thus, a recovery guarantee overcoming the limitations imposed by isotropy and invertibility is desired, which is answered in the next two subsections.

3.2 Tensor Incoherence Conditions

To establish such a guarantee, we need tensor incoherence conditions. First, the tensor bases in the transform domain are introduced:

Definition 3.1 (**Tensor basis**)**.** We denote $\xi_i \in \mathbb{C}^{n_1 \times 1 \times N_3}$ as the tensor column basis with $\xi_i(i,1,:) = 1$ and equalling 0 elsewhere. $\xi_j^{\mathrm{H}} \in \mathbb{C}^{1 \times n_2 \times N_3}$ denotes the tensor row basis with $\xi_j^{\mathrm{H}}(1,j,:) = 1$ and equalling 0 elsewhere. The tensor frontal slice basis is denoted by $\zeta_k \in \mathbb{C}^{1 \times 1 \times n_3}$ with $\zeta_k(1,1,k) = 1$ and equalling 0 elsewhere. Figure 1 provides an illustration.

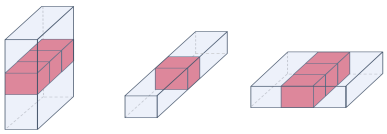


Figure 1: Tensor Basis ξ_i, ζ_k and ξ_j^{H} from Left to Right: The red cubes represent 1 and rest represent 0.

To ensure the energy of the singular tensors is sufficiently spread, we need the following tensor incoherence conditions:

Definition 3.2 (**Tensor incoherence conditions**)**.** Let $T(\mathcal{X}) = \mathcal{U} \star \mathcal{S} \star \mathcal{V}^{\mathrm{H}}$. We say that Program 1 satisfies tensor incoherence condition with parameters $\mu > 0, \nu > 0$ if it holds that

$$\begin{cases} \max\limits_{i} \ \|\mathcal{U}^{\mathrm{H}} \star \xi_i\|_{\mathrm{F}} \le \sqrt{\dfrac{\mu r N_3}{n_1}} \\ \max\limits_{j} \ \|\mathcal{V}^{\mathrm{H}} \star \xi_j\|_{\mathrm{F}} \le \sqrt{\dfrac{\mu r N_3}{n_2}} \end{cases} \tag{2}$$

$$\begin{cases} \max\limits_{ik} \ \|\mathcal{U}^{\mathrm{H}} \star \xi_i \star T(\zeta_k)\|_{\mathrm{F}} \le \sqrt{\dfrac{\nu r}{n_1}} \|\mathbf{T}\|_{1\mapsto 2} \\ \max\limits_{jk} \ \|\mathcal{V}^{\mathrm{H}} \star \xi_j \star T(\zeta_k)\|_{\mathrm{F}} \le \sqrt{\dfrac{\nu r}{n_2}} \|\mathbf{T}\|_{1\mapsto 2} \\ \max\limits_{ik} \ \|\mathcal{U}^{\mathrm{H}} \star \xi_i \star \tilde{T}(\zeta_k)\|_{\mathrm{F}} \le \sqrt{\dfrac{\nu r}{n_1}} \|\tilde{\mathbf{T}}\|_{1\mapsto 2} \\ \max\limits_{jk} \ \|\mathcal{V}^{\mathrm{H}} \star \xi_j \star \tilde{T}(\zeta_k)\|_{\mathrm{F}} \le \sqrt{\dfrac{\nu r}{n_2}} \|\tilde{\mathbf{T}}\|_{1\mapsto 2}, \end{cases} \tag{3}$$

where $\tilde{\mathbf{T}} = (\mathbf{T}^{\dagger})^{\mathrm{H}}$. Parameters μ, ν reflect the energy distribution of the singular tensors of $T(\mathcal{X})$: When

[4] One can verify that the tensor nuclear norm that follows Definition 2.8 is indeed the dual norm of the tensor spectral norm.

the energy of $\mathcal{U}(:,:,k), \mathcal{V}(:,:,k)$ is most uniformly distributed, μ, ν have the minimum values of 1; When $\mathcal{U}(:,:,k)$ or $\mathcal{V}(:,:,k)$) contains an identity matrix, μ, ν have the maximum values of $\frac{n_1}{r}$, $\frac{n_2}{r}$, respectively.

Remark 3.3. When $\mathbf{T}$ is orthogonal, 3 reduces to (14) and (15) in (Lu, 2021). Also (Lu, 2021) states that if letting $\mu = \nu$, 3 can be derived from 2, which means that 2 is stronger and only one set of incoherence conditions is needed. In fact, $\mu = \nu$ does not necessarily hold. Thus for the precision of our model and proof, we keep both incoherence conditions.

3.3 Exact Completion Guarantee

Now we give the exact tensor completion guarantee with arbitrary linear transforms:

Theorem 3.4. *Suppose* $(\mathcal{X}, T)$ *satisfies* 2-3 *with* $\mu, \nu > 0, \lambda = \max(\mu, \nu)$,[5] *then there exist constants* $c_0, c_1, c_2 > 0$ *such that if*

$$p \geq c_0(\gamma^2(\mathbf{T}) + \rho^2(\mathbf{T}))\cdot \frac{\lambda r(n_1+n_2)}{n_1 n_2} \log^2(\gamma(\mathbf{T})(n_1+n_2)N_3), \qquad 4$$

then $\mathcal{X}$ *is the unique optimal solution to* 1 *with probability at least* $1 - c_1((n_1+n_2)N_3)^{-c_2}$, *where* $\gamma(\mathbf{T}) = \max\{\sigma^2_{\max}(\mathbf{T}), 1/\sigma^2_{\min}(\mathbf{T})\}$ *and* $\rho(\mathbf{T}) = N_3 \max(\|\mathbf{T}\|^2_\infty, \|\mathbf{T}^\dagger\|^2_\infty)$. [6]

Remark 3.5 (**Consistency with the existing TNN result**). It can be readily checked that when $\mathbf{T}$ is the DFT matrix[7], 2 is the same as 3. Since for DFT, $N_3 = n_3, \kappa(\mathbf{T}) = 1, \rho(\mathbf{T}) = 1$, Theorem 3.4 regresses to the existing TNN result.

Remark 3.6 (**How to design T according to Theorem 3.4**). From the perspective of minimizing p, if $\kappa(\mathbf{T})$ and $\rho(\mathbf{T})$ take smaller values, then $\mathbf{T}$ is a good measurement matrix (condition number close to 1 and energy uniformly distributed). However, the presence of incoherence parameter λ makes the optimal $\mathbf{T}$ data-dependent and thus coupled in a complex way. Although some experiment results are provided, the design of the optimal transform is outside the scope of this work.

The bound given in Theorem 3.4 is order-wise optimal, neglecting the logarithmic factor, since a rank-r tensor already has $r(n_1+n_2)N_3$ degrees of freedom. Theorem 3.4 is applicable to arbitrary transforms overcoming requirements on isotropy or invertibility. It provides a theoretical guarantee for transforms such as wavelet, Framelet, or data-driven types and greatly enhances the flexibility of the transformed tensor completion model.

3.4 Roadmap and Main Challenges of the Proof

The high-level roadmap of the proof of Theorem 3.4 is to leverage the power of convex optimization. First, a dual certificate is assumed to have been constructed successfully and we need to prove that with this dual certificate, Program 1 has the unique optimal solution and $\mathcal{X}$ is completed exactly. This step is given by Proposition A.1. Second, with the tensor incoherence conditions, it suffices to construct such certificate satisfying subgradient conditions, which is given by Proposition A.3. Combining these two propositions completes the proof of Theorem 3.4.

It is worth noting that the proof of Theorem 3.4 is challenging. Firstly, the proof framework is different. Existing definitions of transform-involved tensor-tensor product, t-SVD and tensor nuclear norm in existing works cannot be utilized in our case, which makes us turn to work directly in the transform domain. Secondly, to the best of our knowledge, the proof of subgradient optimality conditions (Proposition A.1) which requires a case-by-case analysis, is new in the problem of tensor completion. The proof of this proposition is inspired by (Lee et al., 2018), which aims to recover 1-D data sparse in the transform domain. By noticing the analogy among 1-D sparse signal, 2-D low rank matrices and 3-D low rank tensors, we complete the proof of Proposition A.1. Thirdly, the construction of dual certificate requires the establishment of three norm-controlling lemmas (Lemma A.5, A.6, A.7), in which the operator is mingled with the transform $\mathbf{T}$. Bounding Norms such as $\|\mathcal{P}_\mathbb{S} T \mathcal{P}_\Omega T^\dagger \mathcal{P}_\mathbb{S}\|$, $\|T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_\infty$, and $\|\mathcal{P}_\mathbb{S} T(e_{ijk})\|^2_{\mathrm{F}}$ that have not shown up in existing proofs have brought new challenges. Details are in Appendix A.

3.5 Why Slim Transforms Are Better?

Previous works (Jiang et al., 2020, 2023; Luo et al., 2022) give an intuitive explanation that slim transforms can induce redundancy and achieve a better low-rank representation in the transform domain, but no theoretical justification is provided. Here, we present a conditional theorem that theoretically demonstrates the superiority of slim transforms.

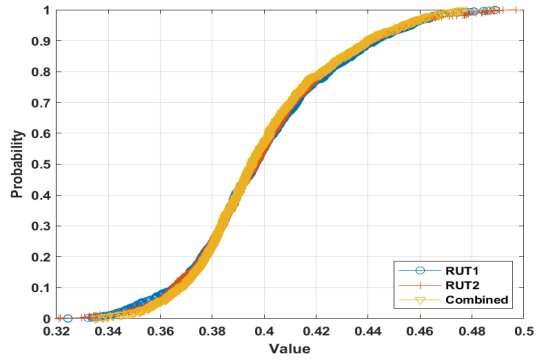
From our proof, it can be observed that the sampling rate p to achieve perfect completion is proportional to two energy terms: $E_1 = \max_{ijk}\{\|\mathcal{P}_\mathbb{S}\mathbf{T}(e_{ijk})\|^2_{\mathrm{F}} \vee \|\mathcal{P}_\mathbb{S}\tilde{\mathbf{T}}(e_{ijk})\|^2_{\mathrm{F}}\}, E_2 = \max_{ijk}\|\mathbf{T}^\dagger\mathcal{P}_\mathbb{S}\mathbf{T}(e_{ijk})\|^2_{\mathrm{F}}$. Let us use a special class of slim transforms, i.e., combinational transforms to see how the value of N_3 impacts

[5] Please note that μ, ν and λ depend on $(\mathcal{X}, T)$.

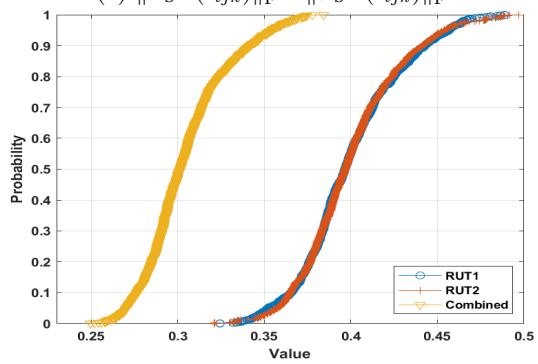
[6] $\gamma(\mathbf{T})$ is a parameter that reflects the spectral distortion of $\mathbf{T}$. A perhaps more familiar related quantity is the condition number $\kappa(\mathbf{T}) = \sigma_{\max}(\mathbf{T})/\sigma_{\min}(\mathbf{T})$. In fact, for any $\mathbf{T}$ under consideration, there exists $c > 0$ such that $\gamma(c\mathbf{T}) = \kappa(\mathbf{T}) = \kappa(c\mathbf{T})$.

[7] The DFT matrix referred to here is with a normalizing parameter $\frac{1}{\sqrt{n_3}}$.

E_1 and E_2.[8]



(a) $\|\mathcal{P}_{\mathbb{S}}\mathbf{T}(e_{ijk})\|_{\mathrm{F}}^2 \vee \|\mathcal{P}_{\mathbb{S}}\tilde{\mathbf{T}}(e_{ijk})\|_{\mathrm{F}}^2$



(b) $\|\mathbf{T}^{\dagger}\mathcal{P}_{\mathbb{S}}\mathbf{T}(e_{ijk})\|_{\mathrm{F}}^2$

Figure 2: Empirical Cumulative Distribution of Two Energy Terms: Note that although the bound is given by $\max\{\cdot\}$, the distribution plays an important role.

First, we analyze E_1. Consider that our slim transform $\mathbf{T}_{\mathrm{c}}$ is formed by concatenating a list of square balanced transforms as $\mathbf{T}_{\mathrm{c}} = \frac{1}{\sqrt{n}}[\mathbf{T}_1^{\mathrm{T}}, \mathbf{T}_2^{\mathrm{T}}, \ldots, \mathbf{T}_M^{\mathrm{T}}]^{\mathrm{T}}$, where the normalizier $\frac{1}{\sqrt{n}}$ is to ensure that the combined $\mathbf{T}_{\mathrm{c}}$ is also balanced. Since the tensor-tensor product is operated slice-wise, the following relations can be acquired: $\forall\ i,j,k,\ \|\mathcal{P}_{\mathbb{S}}\mathbf{T}_c(e_{ijk})\|_{\mathrm{F}}^2 = \mathrm{mean}\{\|\mathcal{P}_{\mathbb{S}}\mathbf{T}_m(e_{ijk})\|_{\mathrm{F}}^2\}$, $\|\mathcal{P}_{\mathbb{S}}\tilde{\mathbf{T}}_{\mathrm{c}}(e_{ijk})\|_{\mathrm{F}}^2 = \mathrm{mean}\{\|\mathcal{P}_{\mathbb{S}}\tilde{\mathbf{T}}_m(e_{ijk})\|_{\mathrm{F}}^2\}$, which means that the two norms in E_1 of the slim transform $\mathbf{T}_c$ are averaged by those of its components. After taking the maximum w.r.t. i,j,k, it becomes an inequality: $E_1(\mathbf{T}_c) \leq \mathrm{mean}\{E_1(\mathbf{T}_m)\}$.[9]

However, the relations become different when analyzing E_2 due to $\mathbf{T}_{\mathrm{c}}{}^{\dagger}\mathcal{P}_{\mathbb{S}}\mathbf{T}_{\mathrm{c}}(e_{ijk})$. One observes that: $\forall\ i,j,k,\ \mathbf{T}_{\mathrm{c}}^{\dagger}\mathcal{P}_{\mathbb{S}}\mathbf{T}_{\mathrm{c}}(e_{ijk}) = \mathrm{mean}(\mathbf{T}_m^{\dagger}\mathcal{P}_{\mathbb{S}}\mathbf{T}_m(e_{ijk}))$, which means, instead of the energy, the (entries of) tensors themselves are the mean of the components. If we assume entries in $\mathbf{T}_m^{\dagger}\mathcal{P}_{\mathbb{S}}\mathbf{T}_m(e_{ijk})$ are i.i.d., Central Limit Theorem implies that the energy of $\mathbf{T}_m^{\dagger}\mathcal{P}_{\mathbb{S}}\mathbf{T}_m(e_{ijk})$ is shrunken compared to the energy of each of its components, granting slim transforms the advantage over square ones as stated in the following theorem:

Theorem 3.7. *If the sampling rate p satisfies 87, 89, 92, 93, then the application of the combinational slim transform $\mathbf{T}_c$ requires a smaller minimum sampling rate p than square transforms $\mathbf{T}_m$. Specifically, denoting the minimum sampling rate using $\mathbf{T}_c$ as p_c and the minimum sampling rate using $\mathbf{T}_m$ as p_m, we have*

$$\frac{p_c}{p_m} \sim \mathcal{O}\left(\frac{E_2(\mathbf{T}_c)}{E_2(\mathbf{T}_m)}\right). \tag{5}$$

It is challenging to unconditionally remove the assumptions in the above theorem. Nonetheless, these assumptions are validated by experiments shown in Figure 2 (with additional results in Figure 11). In particular, the energy terms in 87, 89, 92, 93, are all of E_1-type, which are not proportional to the minimum sampling rate p_c and p_m according to Figure 2 and Figure 11. Therefore, 87, 89, 92, 93, are already satisfied near the minimum sampling rate. The proof of Theorem 3.7 is provided in Appendix E.

The ratio p_c/p_m in the above theorem quantifies the improvement due to using slim transforms. This ratio corresponds to a decrease in the sampling rate by a factor close to n_3/N_3.

4 EXPERIMENTS

Numerical experiments are first conducted on random data to study the completion phenomena and then application to visual data impainting is considered.[10].

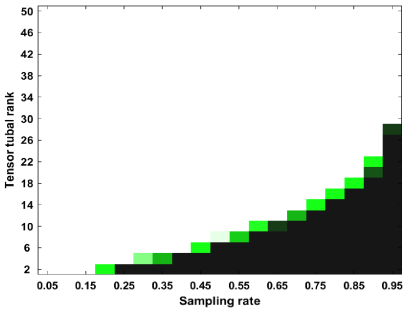
4.1 Exact Completion of Random Tensors

First, we examine the completion behaviors of Program 1 with $\mathrm{rank}(\mathcal{M})$ and sampling rate p both varying, to support Theorem 3.4. To that goal, we need to generate random tensors $T(\mathcal{M})$ with specified tubal rank r. Such generation of $\mathcal{M}$ is not that trivial since the randomly generated $T(\mathcal{M}) \in \mathbb{C}^{n_1\times n_2\times N_3}$ with a tubal rank might not have preimage $\mathcal{M} \in \mathbb{C}^{n_1\times n_2\times n_3}$ when $T(\cdot)$ is not bijection ($N_3 > n_3$). Thus, to generate random tensors $\mathcal{M}$ satisfying $T(\mathcal{M}) = r$, we design an alternating projection algorithm to iteratively project a randomly initialized tensor onto the target manifold. Elaborate explanation and corresponding algorithm (Algorithm 4) are provided in Appendix G. Using Algorithm 4, we generate $\mathcal{M} \in \mathbb{C}^{n_1\times n_2\times n_3}$ with $n_1 = n_2 = 50, n_3 = 20$ and $N_3 = \{20, 40\}$. We have tested two types of $\mathbf{T}$: DFT and random unitary transform (RUT), using the first n_3 columns when $N_3 > n_3$.

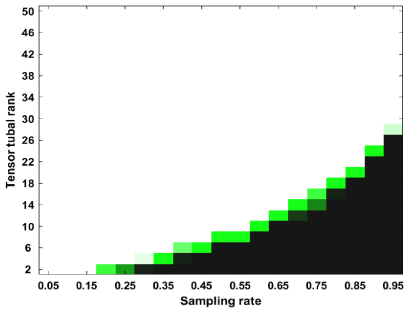
[8] There are other terms that affect p. The formulations and analysis of them are similar to E_1 and E_2. Please find a complete analysis in Appendix E.

[9] This inequality will be observed in Figure 11a in Appendix E.

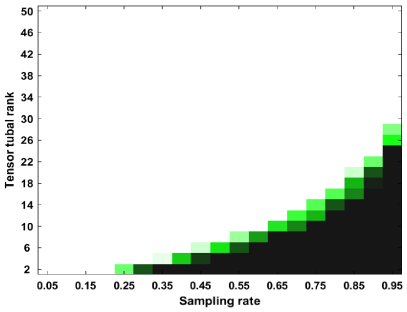
[10] The source codes can be found at https://github.com/vecevecev/transformed_TNN



(a) Comparing Performances of (1) Orthogonal Transform (DFT, $N_3 = n_3$), and (2) Slim Transform (DFT, $N_3 = 2n_3$)

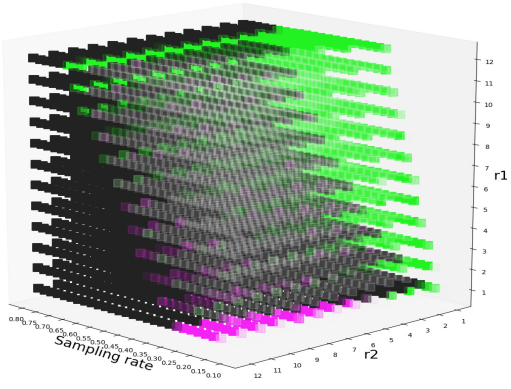


(b) Comparing Performances of (1) Orthogonal Transform (RUT, $N_3 = n_3$), and (2) Slim Transform (RUT, $N_3 = 2n_3$)

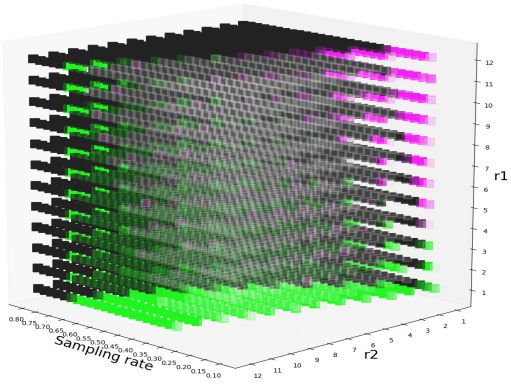


(c) Comparing Performances of (1) Slim Transform (RUT, $\gamma(\mathbf{T}) \approx 4$), and (2) Slim Transform (RUT, $\gamma(\mathbf{T}) = 1$)

Figure 3: Exact Completion with Varying Rank (x-axis) and Sampling Rate (y-axis): The white cube denotes all failure and the green/black cube denotes all success (in 10 trials of Monte Carlo experiments). The black cubes stand for overlapping performance region of (1) and (2), and the green cubes stand for the performance gain of (2) over (1) in each subfigure.



(a) Comparing Performances of (1) Orthogonal Transform (RUT1, $N_3 = n_3$), and (2) Slim Transform (RUT1-RUT2 Combined, $N_3 = 2n_3$)



(b) Comparing Performances of (1) Orthogonal Transform (RUT2, $N_3 = n_3$), and (2) Slim Transform (RUT1-RUT2 Combined, $N_3 = 2n_3$)

Figure 4: Exact Completion with Varying Rank($T_1(\mathcal{X})$) (z-axis), Rank($T_2(\mathcal{X})$) (y-axis) and Sampling Rate (x-axis): Black cubes stand for overlapping parts. Green cubes stand for the performance gain of the combined transform over RUT1 and RUT2. Magenta cubes stand for the region that RUT1 or RUT2 outperforms the combined transform.

We have also conducted experiments to see the influence of $\kappa(\mathbf{T})$, in which the singular values of the generated RUT are randomly set between $[0.5, 2]$, resulting $\gamma(\mathbf{T})(\kappa(\mathbf{T})) \approx 4$[11]. The sampling rate (SR) is set as $p = [0.05 : 0.05 : 0.95]$ and $\text{rank}(\mathcal{M}) = [2 : 2 : 50]$. We conducted 10 test cases for each (r, p) and in each case if $\|\mathcal{X} - \mathcal{M}\|_{\text{F}}/\|\mathcal{X}\|_{\text{F}} \leq 10^{-3}$ we deem $\mathcal{X}$ completed successfully. The parameters are set as $\alpha_0 = 10^{-2}, \alpha_{\max} = 10^6, \rho = 1.02$ and the ADMM algorithm terminates when $\max(\|\mathcal{X}^t - \mathcal{X}^{t-1}\|_\infty, \|\mathcal{Y}^t - \mathcal{Y}^{t-1}\|_\infty, \|T(\mathcal{X}^t) - T(\mathcal{X}^{t-1})\|_\infty) \leq 10^{-10}$.

The performances of each transform are provided in Figure 6 in Appendix D. Here, we superimpose the first and the second row of Figure 6 for clearer comparison, as shown in Figure 3. From Figure 6, it can be observed that the completion behaviors are alike among different types of $\mathbf{T}$ and different N_3, all showing a region of exact completion in all cases which supports Theorem 3.4. From Figure 3a and Figure 3b, it can also be observed that, with the same tubal ranks, slim transforms that map $\mathcal{X}$ into higher dimensional space outperforms their counterparts, which validates the advantage of such non-invertible transforms. Additionally, from Figure 3c, the transform with $\gamma(\mathbf{T})(\kappa(\mathbf{T})) = 1$ exhibits better recovery efficacy than the transform with $\gamma(\mathbf{T}) \approx 4$, also supporting Theorem 3.4.

Such non-invertible slim transforms can provide more information since more prior knowledge of $\mathcal{X}$ is available. To illustrate this, consider two transforms $\mathbf{T}_1$ and $\mathbf{T}_2$. We generate $\mathcal{X}$ by Algorithm 4 to satisfy $\text{rank}(T_1(\mathcal{X})) = r_1$ and $\text{rank}(T_2(\mathcal{X})) = r_2$. Since $\mathbf{T}_1$ and $\mathbf{T}_2$ provide information of $\mathcal{X}$ from different domains, the concatenated slim transform $\mathbf{T}_3 =$

[11] Since the singular values are random from $[0.5, 2]$, in some test cases $\kappa(\mathbf{T})$ and $\gamma(\mathbf{T})$ are a bit smaller than 4.

$[\mathbf{T}_1^{\mathrm{T}}, \mathbf{T}_2^{\mathrm{T}}]^{\mathrm{T}}$ ensembles richer information than its individual components. We set $\mathbf{T}_1$ and $\mathbf{T}_2$ as two RUTs, the slim $\mathbf{T}_3$ is cascaded as $\mathbf{T}_3 = \frac{1}{\sqrt{2}}[\mathbf{T}_1^{\mathrm{T}}, \mathbf{T}_2^{\mathrm{T}}]^{\mathrm{T}}$ with $r_1, r_2 = [1 : 12]$, $p = [0.1 : 0.025 : 0.8]$ and plot 3D phase graphs shown in Figure 4.

It can be observed from Figure 4 (the unimposed version is included in Appendix D as Figure 7) that: 1) As expected, the performance of $\mathbf{T}_1$ ($\mathbf{T}_2$) basically remains fixed with r_2 (r_1) varying; 2) As shown in Figure 4a and Figure 4b, though the tubal rank of $T_3(\mathcal{X})$ has not decreased ($\mathrm{rank}(T_3(\mathcal{X})) = \mathrm{rank}(T_1(\mathcal{X})) \vee \mathrm{rank}(T_2(\mathcal{X}))$), compared to using $\mathbf{T}_1/\mathbf{T}_2$ only, the exact completion region is enlarged giving credit to the utilization of the additional information provided by another transform. Such a simple and natural combinational design of transform $T(\cdot)$ provides a perspective of understanding why slim transforms may have better performances.

4.2 Application to Visual Data Inpainting

We conduct experiments on visual data. Since the experimental superiorty of slim transforms have been sufficiently validated in works like (Jiang et al., 2020; Luo et al., 2022), we focus on studying the impacts of different transforms and have not included methods with different tensor structures and assumptions, such as CP, Tucker, tensor ring, etc.. Readers can refer to these works for such comparisons. Additional experiment results are in Appendix D.

The transforms used for comparison include $n_3 \times n_3$ DFT, i.e., TNN, $n_3 \times n_3$ DCT, the first n_3 columns of $N_3 \times N_3$ DFT matrix and DCT matrix with $N_3 = 2n_3$ (DFT2, DCT2), DFT/DCT cantencated transform (DFT-DCT), discrete Walsh-Hadamard transform (DWHT) and Framelet (FTNN) (Jiang et al., 2020) with piecewise cubic B-spline and decomposition level $l = 4$. The ADMM parameters are $\alpha_0 = 10^{-2}, \alpha_{\max} = 10^6, \rho = 1.2$ and the algorithm terminates when $\max(\|\mathcal{X}^t - \mathcal{X}^{t-1}\|_\infty, \|\mathcal{Y}^t - \mathcal{Y}^{t-1}\|_\infty, \|T(\mathcal{X}^t) - T(\mathcal{X}^{t-1})\|_\infty) \leq 10^{-3}$ (10^{-4} for Framelet).

4.2.1 MSI Data

We first apply our model on MSI data, of which the third dimension contains rich spectrum information. Each MSI is of size $512 \times 512 \times 31$ (height $\times$ width $\times$ band). The methods are tested on MSI "super balls", "cloth", "stuffed toys", "photo and face", "oil painting" and "clay". The evaluation metrics adopted are the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM).

The top half of Table 1 shows the average PSNR and SSIM value on MSI data for SR= $0.05, 0.1, 0.15$. [12]

[12] The experiment results are stable, shown by statistical test results in Appendix D.

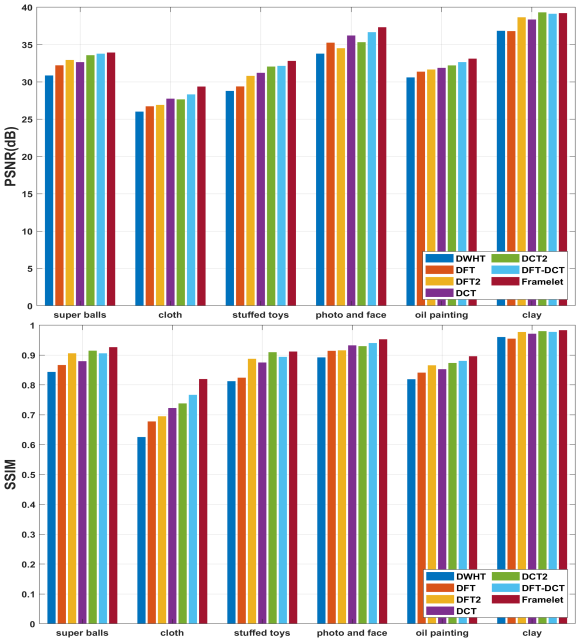


Figure 5: The PSNR (left) and SSIM (right) results on MSI Data with SR=0.05.

Figure 5 lists the indicators for each MSI. It is observable that the non-invertible slim transforms, i.e., DFT2 and DCT2, outperform their square counterparts from which they originate. The Framelet transform and DFT-DCT achieve the best and the second best results, both being slim (the Framelet transform is slimmer as we set $l = 4$). It is worth noting that the recovery performance of DFT-DCT is quite satisfactory by further considering that such design is simple and relatively efficient (the dimension is only doubled after the transform). Naturally more transforms could be included and each of them provides information about the original data from a different domain, which reflects the idea of ensemble learning or information fusion. Please find more experiment results in Appendix D (indicators in Figure 5, visual recovery results in Figure 9).

4.2.2 Video Data

In this experiment, the application of video data recovery is considered. The momentum of applying tensor-based modeling is that the frames of video data are correlated and can be viewed as the third dimension of tensors. The test videos include "bus", "carphone", "foreman", "miss", "coastguard", "news", "soccer", "suzie" and "trevor". Contents of these videos contain cars, boats, rivers, and humans of different resolutions. Also, both static and dynamic scenes occur in these test videos. The assessment criteria are the mean PSNR and the mean SSIM of all frames.

The average mPSNR and mSSIM values are given by

Table 1: The Average Performances on MSI And Video Data: The best and the second best results are highlighted by bolder and italic fonts, respectively.

Data	SR	DWHT	DFT	DFT2	DCT	DCT2	DFT-DCT	Framelet
MSI	0.05	31.147/0.826	31.959/0.847	32.578/0.875	33.008/0.872	33.352/0.891	*33.777/0.894*	**34.292/0.915**
	0.1	35.683/0.919	37.021/0.937	37.453/0.949	38.153/0.951	38.228/0.956	*38.793/0.959*	**39.192/0.968**
	0.15	38.175/0.948	39.774/0.963	40.243/0.971	40.942/0.972	40.902/0.974	*41.488/0.976*	**41.889/0.981**
Video	0.05	23.168/0.605	23.147/0.585	23.832/0.650	23.392/0.595	*24.166/0.661*	24.086/0.653	**24.539/0.701**
	0.1	25.733/0.710	25.738/0.686	26.607/0.753	25.960/0.694	*26.839/0.757*	26.743/0.749	**27.226/0.787**
	0.15	27.087/0.760	27.190/0.741	28.126/0.802	27.404/0.747	*28.297/0.803*	28.198/0.797	**28.715/0.831**

the bottom half of Table 1 and specifically, the performance indicators of each test video for SR = 0.05 are shown in Figure 8 in Appendix D. As can be observed, the non-invertible slim transforms exhibit superior performance among the videos containing different contents and scenes, with a more obvious advantage in mSSIM. More experiment results can be found in Appendix D (Figure 8 and 10).

4.3 On the Performance-Cost Trade-off of the Slim Transform

Here we provide a trade-off analysis on the slimness parameter, i.e., N_3/n_3. Testing on MSI "stuffed toys", we fix the transform type to be Framelet and change the level parameter to yield ablation on N_3. For each N_3/n_3, we record the performances as well as the wall-clock time and memory usage. The result is displayed in Table 2.

Table 2: Trade-off Analysis.

Slimness (N_3/n_3)	3	6	9	12
PSNR (dB)	25.85	30.48	33.77	32.81
SSIM	0.756	0.876	0.913	0.902
Wall-clock time (sec)	282.0	816.1	1248.0	1677.7
Peak memory usage (MB)	745.7	1490.9	2236.4	2981.8

As observed in Table 2, the performance improves with N_3/n_3 until $N_3/n_3 = 12$, after which it drops due to the increase introduced by the logarithmic factor in the sampling-rate bound. The wall-clock running time and memory usage grow almost linearly with N_3, which is consistent with the complexity analysis in Appendix H of the manuscript. Specifically, the computational complexity is $O\left(\max(n_1, n_2) \cdot \min(n_1, n_2)^2 N_3 + n_1 n_2 n_3 N_3\right)$ per-iteration for all transforms, and is therefore linear in N_3. Thus, choosing N_3/n_3 within the range $[6, 12]$ provides a good balance in this tradeoff.

On the other hand, the dominant cost arises from the t-SVD operation, whose per-iteration complexity is $O\left(\max(n_1, n_2) \cdot \min(n_1, n_2)^2 N_3\right)$. In practice, this step is highly parallelizable: SVDs of all N_3 slices can be computed simultaneously since they are independent. As a result, the additional running-time overhead introduced by slim transforms can be largely mitigated.

5 CONCLUSION

In this work, the theoretical guarantee of exact tensor completion with arbitrary linear transforms is established by directly operating the tensors in the transform domain. Breaking the constraints of isotropy and invertibility, our model and proof greatly enhance the flexibility of tensor completion. Moreover, we discloses why slim transforms outperform their square counterparts theoretically, supporting recent emerging experimental works. Extensive experiments validate the superiority of the proposed method.

6 ACKNOWLEDGMENTS

Li would like to express special thanks to Prof. Jiang and the members of her group for making his Master's experience truly memorable. He is deeply grateful to the friends, and the people he met at SJTU, who have brought warmth and light to those days and filled them with memories.

References

J. A. Bengua, H. N. Phien, H. D. Tuan, and M. N. Do. Efficient tensor completion for color image and video recovery: Low-rank tensor train. *IEEE Transactions on Image Processing*, 26(5):2466–2479, 2017.

E. J. Candès and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust

principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.

Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis. Low-rank matrix recovery from errors and erasures. *IEEE Transactions on Information Theory*, 59(7): 4324–4337, 2013.

L. De Lathauwer and B. De Moor. From matrix to tensor: Multilinear algebra and signal processing. In *Institute of mathematics and its applications conference series*, volume 67, pages 1–16. Citeseer, 1998.

D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011. doi: 10.1109/TIT.2011.2104999.

C. He, Y. Xu, Z. Wu, S. Zheng, and Z. Wei. Multi-dimensional visual data restoration: Uncovering the global discrepancy in transformed high-order tensor singular values. *IEEE Transactions on Image Processing*, 33:6409–6424, 2024. doi: 10.1109/TIP.2024.3475738.

T.-X. Jiang, M. K. Ng, X.-L. Zhao, and T.-Z. Huang. Framelet representation of tensor nuclear norm for third-order tensor completion. *IEEE Transactions on Image Processing*, 29:7233–7244, 2020.

T.-X. Jiang, X.-L. Zhao, H. Zhang, and M. Ng. Dictionary learning with low-rank coding coefficients for tensor completion. *IEEE Transactions on Neural Networks and Learning Systems*, 34(2):932–946, 2023.

E. Kernfeld, M. Kilmer, and S. Aeron. Tensor–tensor products with invertible linear transforms. *Linear Algebra and its Applications*, 485:545–570, 2015. ISSN 0024-3795. doi: https://doi.org/10.1016/j.laa.2015.07.021.

T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

P. Korohoda and A. Dąbrowski. Generalized convolution concept based on dct. In *2004 12th European Signal Processing Conference*, pages 973–976, 2004.

K. Lee, Y. Li, K. H. Jin, and J. C. Ye. Unified theory for recovery of sparse signals in a general transform domain. *IEEE Transactions on Information Theory*, 64(8):5457–5477, 2018.

Y. Li, W. Liang, K. Xie, D. Zhang, S. Xie, and K. Li. Lightnestle: Quick and accurate neural sequential tensor completion via meta learning. In *IEEE INFOCOM 2023 - IEEE Conference on Computer Communications*, pages 1–10, 2023.

J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):208–220, 2013.

C. Lu. Transforms based tensor robust pca: Corrupted low-rank tensors recovery via convex optimization. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1125–1132, 2021.

C. Lu, X. Peng, and Y. Wei. Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5989–5997, 2019.

C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4): 925–938, 2020.

Y.-S. Luo, X.-L. Zhao, T.-X. Jiang, Y. Chang, M. K. Ng, and C. Li. Self-supervised nonlinear transform-based tensor nuclear norm for multi-dimensional image recovery. *IEEE Transactions on Image Processing*, 31:3793–3808, 2022.

C. Mu, B. Huang, J. Wright, and D. Goldfarb. Square deal: Lower bounds and improved relaxations for tensor recovery. In E. P. Xing and T. Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 73–81, Bejing, China, 22–24 Jun 2014. PMLR.

I. V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

Y. Panagakis, J. Kossaifi, G. G. Chrysos, J. Oldfield, M. A. Nicolaou, A. Anandkumar, and S. Zafeiriou. Tensor methods in computer vision and deep learning. *Proceedings of the IEEE*, 109(5):863–890, 2021.

W. Qin, H. Wang, F. Zhang, J. Wang, X. Luo, and T. Huang. Low-rank high-order tensor completion with applications in visual data. *IEEE Transactions on Image Processing*, 31:2433–2448, 2022.

N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65 (13):3551–3582, 2017.

Q. Song, H. Ge, J. Caverlee, and X. Hu. Tensor completion algorithms in big data analytics. *ACM Trans. Knowl. Discov. Data*, 13(1), 2019. ISSN 1556-4681. doi: 10.1145/3278607.

R. Stephan, S. Oleksandr, and G. Stephan. Introduction to tensor decompositions and their applications in machine learning, 2017.

J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12:389–434, 2012.

J.-L. Wang, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, and M. K. Ng. Multi-dimensional visual data completion via low-rank tensor representation under coupled transform. *IEEE Transactions on Image Processing*, 30:3581–3596, 2021. doi: 10.1109/TIP.2021.3062995.

H. Yin. Tensor sparse representation for 3-d medical image fusion using weighted average rule. *IEEE Transactions on Biomedical Engineering*, 65(11): 2622–2633, 2018.

L. Yuan, C. Li, D. Mandic, J. Cao, and Q. Zhao. Tensor ring decomposition with rank minimization on latent space: An efficient approach for tensor completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 9151–9158, 2019.

Z. Zhang and S. Aeron. Exact tensor completion using t-svd. *IEEE Transactions on Signal Processing*, 65 (6):1511–1526, 2017.

L. Zhao, X. Xu, Z. Wang, Y. Zhang, B. Zhang, W. Zheng, D. Du, J. Zhou, and J. Lu. Lowrankocc: Tensor decomposition and low-rank recovery for vision-based 3d semantic occupancy prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9806–9815, June 2024.

Q. Zhao, L. Zhang, and A. Cichocki. Bayesian cp factorization of incomplete tensors with automatic rank determination. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1751–1763, 2015.

Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki. Tensor ring decomposition. *arXiv preprint arXiv:1606.05535*, 2016.

Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, Q. Zhao, and T.-X. Jiang. Fully-connected tensor network decomposition and its application to higher-order tensor completion. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12):11071–11078, May 2021. doi: 10.1609/aaai.v35i12.17321.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes. See Appendix H.]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes. See Table 3 in Appendix D.]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes]
   (b) The license information of the assets, if applicable. [Yes]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Yes]
   (d) Information about consent from data providers/curators. [Yes]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Yes]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. Not Applicable]

Appendices

The appendices are organized as follows: Appendix A contains the proof of the main theorem. Appendix B proves three crucial lemmas and the energy terms are bounded in Appendix C. Additional experiment results are in Appendix D. The complete analysis of the sampling rate is given in Appendix E. The algorithm for solving Program 1 is provided in Appendix F. Methods of generating random tensors in Section 4.1 are given in Appendix G. Limitations are in Appendix H.

A Proof of Theorem 3.4

The proof of Theorem 3.4 consists of the proofs of two propositions: the optimality condition (Proposition A.1) and construct dual certificate (Proposition A.3). The latter further requires three norm bounding lemmas and bounding the energy terms, which are contained in Appendix B and Appendix C.

A.1 Optimality Condition

Proposition A.1. *Tensor $\mathcal{X}$ is the unique optimal solution to 1 if the following conditions hold:*

- $$\|\mathcal{P}_{\mathbb{S}} T\mathcal{R}_{\Omega} T^{\dagger}\mathcal{P}_{\mathbb{S}} - \mathcal{P}_{\mathbb{S}} TT^{\dagger}\mathcal{P}_{\mathbb{S}}\| \leq \frac{1}{2} \tag{6}$$

  where $\mathcal{R}_{\Omega} = \frac{1}{p}\mathcal{S}_{\Omega}$, $\mathbb{S}$ denotes the subspace spanned by singular tensors of $T(\mathcal{X}) = \mathcal{U} \star \mathcal{S} \star \mathcal{V}^H$:

$$\begin{aligned}
&\mathbb{S} = \{\mathcal{U} \star {\Gamma_1}^H + \Gamma_2 \star \mathcal{V}^H \mid \forall\ \Gamma_1 \in \mathbb{C}^{n_2 \times r \times N_3}, \Gamma_2 \in \mathbb{C}^{n_1 \times r \times N_3}\} \\
&\mathcal{P}_{\mathbb{S}}(\mathcal{Z}) = \mathcal{U} \star \mathcal{U}^H \star \mathcal{Z} + \mathcal{Z} \star \mathcal{V} \star \mathcal{V}^H - \mathcal{U} \star \mathcal{U}^H \star \mathcal{Z} \star \mathcal{V} \star \mathcal{V}^H \\
&\mathcal{P}_{\mathbb{S}^{\perp}}(\mathcal{Z}) = \mathcal{Z} - \mathcal{P}_{\mathbb{S}}(\mathcal{Z}) = (\mathcal{I} - \mathcal{U} \star \mathcal{U}^H) \star \mathcal{Z} \star (\mathcal{I} - \mathcal{V} \star \mathcal{V}^H)
\end{aligned}$$

- $\exists\ \mathcal{Y} \in \mathbb{C}^{n_1 \times n_2 \times N_3}$, *s.t.*

$$(TT^{\dagger} - T\mathcal{S}_{\Omega} T^{\dagger})^H \mathcal{Y} = \mathbf{0} \tag{7}$$

$$\|\mathcal{P}_{\mathbb{S}}(\mathcal{Y}) - \mathcal{U} \star \mathcal{V}^H\|_F \leq \frac{p}{8\kappa(\mathbf{T})} \tag{8}$$

$$\|\mathcal{P}_{\mathbb{S}^{\perp}}(\mathcal{Y})\| \leq \frac{1}{2} \tag{9}$$

A.2 Subgradient

Recall A.2 (Subgradient).

1. By convex optimization theory, $\mathcal{A}$ is in the subgradient set of $\|\cdot\|_*$ defined upon $\mathcal{X}_0$ iff $\forall \mathcal{X}$,

$$\|\mathcal{X}\|_* \geq \|\mathcal{X}_0\|_* + \Re\langle \mathcal{A}, \mathcal{X} - \mathcal{X}_0 \rangle.$$

2. the subgradient formulation of $\|\cdot\|_*$ is given as

$$\mathcal{A} \in \{\mathcal{U} \star \mathcal{V}^{\mathrm{H}} + \mathcal{W} \mid \mathcal{U}^{\mathrm{H}} \star \mathcal{W} = \mathbf{0}, \mathcal{W} \star \mathcal{V} = \mathbf{0}, \|\mathcal{W}\| \leq 1\},$$

   where $\mathcal{U}, \mathcal{V}$ are the singular tensors of $T(\mathcal{X})$.

A.3 Proof of Proposition A.1

Proof. The proof of this proposition is inspired by (Lee et al., 2018), which aims to recover 1-D data sparse in the transform domain. By noticing the analogy among 1-D sparse signal, 2-D low rank matrices and 3-D low rank tensors, the analysis in (Lee et al., 2018) can serve as guidance.

First let us revisit the definition of subgradient in convex optimization and give the subgradient of the natural tensor nuclear norm as Recall A.2.

Suppose $\mathcal{H}$ is a perturbation satisfying $\mathcal{S}_\Omega(\mathcal{H}) = \mathbf{0}$, according to the fact that $\|\mathbf{U}_\perp \star \mathbf{V}_\perp^{\mathrm{H}}\| = 1$, we have

$$\|T(\mathcal{X}+\mathcal{H})\|_* \geq \|T(\mathcal{X})\|_* + \Re\langle \mathcal{U}\star\mathcal{V}^{\mathrm{H}} + \mathcal{U}_\perp \star \mathcal{V}_\perp^{\mathrm{H}}, T(\mathcal{H})\rangle, \tag{10}$$

where $\mathcal{U},\mathcal{V}$ are the singular tensors of $T(\mathcal{X})$, $\mathbb{S}$ is spanned by $\mathcal{U},\mathcal{V}$ and $\mathcal{U}_\perp,\mathcal{V}_\perp$ are the singular tensors of $\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))$.

Now continue to analyze the second term in the RHS of 10 and decompose it into three parts,

$$\langle \mathcal{U}\star\mathcal{V}^{\mathrm{H}} + \mathcal{U}_\perp\star\mathcal{V}_\perp^{\mathrm{H}}, T(\mathcal{H})\rangle = \langle \mathcal{Y}, T(\mathcal{H})\rangle + \langle \mathcal{U}_\perp\star\mathcal{V}_\perp^{\mathrm{H}}, T(\mathcal{H})\rangle - \langle \mathcal{Y} - \mathcal{U}\star\mathcal{V}^{\mathrm{H}}, T(\mathcal{H})\rangle. \tag{11}$$

- First look into $\langle \mathcal{Y}, T(\mathcal{H})\rangle$.

$$\begin{aligned}&\langle \mathcal{Y}, T(\mathcal{H})\rangle\\ &= \underbrace{\langle \mathcal{Y}, (\mathcal{I} - TT^\dagger)T(\mathcal{H})\rangle}_{A_1} + \underbrace{\langle \mathcal{Y}, (TT^\dagger - T\mathcal{S}_\Omega T^\dagger)T(\mathcal{H})\rangle}_{A_2} + \underbrace{\langle \mathcal{Y}, T\mathcal{S}_\Omega T^\dagger T(\mathcal{H})\rangle}_{A_3},\end{aligned} \tag{12}$$

  which can also be trisected as:

$$A_1 = \langle \mathcal{Y}, (\mathcal{I} - TT^\dagger)T(\mathcal{H})\rangle = \langle \mathcal{Y}, T(\mathcal{H}) - T(\mathcal{H})\rangle = 0 \text{ (by } \mathbf{T}^\dagger\mathbf{T} = \mathbf{I}), \tag{13}$$

$$A_2 = \langle \mathcal{Y}, (TT^\dagger - T\mathcal{S}_\Omega T^\dagger)T(\mathcal{H})\rangle = \langle (TT^\dagger - T\mathcal{S}_\Omega T^\dagger)^{\mathrm{H}}\mathcal{Y}, T(\mathcal{H})\rangle = 0 \text{ (by 7)}, \tag{14}$$

$$A_3 = \langle \mathcal{Y}, T\mathcal{S}_\Omega T^\dagger T(\mathcal{H})\rangle = \langle \mathcal{Y}, T\mathcal{S}_\Omega(\mathcal{H})\rangle = 0 \text{ (by } \mathcal{S}_\Omega(\mathcal{H}) = 0), \tag{15}$$

  which combines for $\langle \mathcal{Y}, T(\mathcal{H})\rangle = 0$.

- For $\langle \mathcal{U}_\perp\star\mathcal{V}_\perp^{\mathrm{H}}, T(\mathcal{H})\rangle$, we have

$$\langle \mathcal{U}_\perp\star\mathcal{V}_\perp^{\mathrm{H}}, T(\mathcal{H})\rangle = \langle \mathcal{U}_\perp\star\mathcal{V}_\perp^{\mathrm{H}}, \mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\rangle = \|\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\|_* \tag{16}$$

- For $\langle \mathcal{Y} - \mathcal{U}\star\mathcal{V}^{\mathrm{H}}, T(\mathcal{H})\rangle$, an upper bound can be given:

$$\begin{aligned}&\langle \mathcal{Y} - \mathcal{U}\star\mathcal{V}^{\mathrm{H}}, T(\mathcal{H})\rangle\\ =&\langle \mathcal{P}_{\mathbb{S}}(\mathcal{Y} - \mathcal{U}\star\mathcal{V}^{\mathrm{H}}), T(\mathcal{H})\rangle + \langle \mathcal{P}_{\mathbb{S}^\perp}(\mathcal{Y}), T(\mathcal{H})\rangle - \langle \mathcal{P}_{\mathbb{S}^\perp}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}}), T(\mathcal{H})\rangle\\ \leq&\|\mathcal{P}_{\mathbb{S}}(\mathcal{Y} - \mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_{\mathrm{F}}\cdot\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}} + \|\mathcal{P}_{\mathbb{S}^\perp}(\mathcal{Y})\|\cdot\|\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\|_*.\end{aligned} \tag{17}$$

  In the last equation of 17, we use the Cauchy–Schwarz inequality and the dual definition of tensor spectral norm 2.7 and tensor nuclear norm 2.8.

  Plug 8 and 9 into 17, which yields

$$\Re\langle \mathcal{Y} - \mathcal{U}\star\mathcal{V}^{\mathrm{H}}, T(\mathcal{H})\rangle \leq \frac{p}{8\kappa(\mathbf{T})}\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}} + \frac{1}{2}\|\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\|_* \tag{18}$$

Now 11 can be bounded as

$$\Re\langle \mathcal{U}\star\mathcal{V}^{\mathrm{H}} + \mathcal{U}_\perp\star\mathcal{V}_\perp^{\mathrm{H}}, T(\mathcal{H})\rangle \geq \frac{1}{2}\|\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\|_* - \frac{p}{8\kappa(\mathbf{T})}\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}}. \tag{19}$$

Here the proof is furthered by a case-by-case analysis:

1. case 1: $\frac{1}{2}\|\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\|_* > \frac{p}{8\kappa(\mathbf{T})}\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}}$

   In this case,

$$\begin{aligned}\|T(\mathcal{X}+\mathcal{H})\|_* - \|T(\mathcal{X})\|_* &\geq \Re\langle \mathcal{U}\star\mathcal{V}^{\mathrm{H}} + \mathcal{U}_\perp\star\mathcal{V}_\perp^{\mathrm{H}}, T(\mathcal{H})\rangle\\ &\geq \frac{1}{2}\|\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\|_* - \frac{p}{8\kappa(\mathbf{T})}\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}} > 0\end{aligned}$$

2. case 2: $\frac{1}{2}\|\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\|_* \le \frac{p}{8\kappa(\mathbf{T})}\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}}$

We show that in this complementary case, $\mathcal{H} = \mathbf{0}$. First, $\mathcal{S}_\Omega(\mathcal{H}) = 0 \Rightarrow (\mathcal{S}_\Omega T^\dagger T)(\mathcal{H}) = 0 \Rightarrow (T\mathcal{R}_\Omega T^\dagger + \mathcal{I} - TT^\dagger)T(\mathcal{H}) = 0$. Thus we have

$$0 = \underbrace{\langle \mathcal{P}_{\mathbb{S}}(T(\mathcal{H})), (T\mathcal{R}_\Omega T^\dagger + \mathcal{I} - TT^\dagger)\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\rangle}_{B_1} + \underbrace{\langle \mathcal{P}_{\mathbb{S}}(T(\mathcal{H})), (T\mathcal{R}_\Omega T^\dagger - TT^\dagger)\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\rangle}_{B_2}.$$

$$\begin{aligned} |B_1| &= |\langle \mathcal{P}_{\mathbb{S}}(T(\mathcal{H})), \mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\rangle + \langle \mathcal{P}_{\mathbb{S}}(T(\mathcal{H})), (T\mathcal{R}_\Omega T^\dagger - TT^\dagger)\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))| \\ &\ge (1 - \|\mathcal{P}_{\mathbb{S}} T\mathcal{R}_\Omega T^\dagger \mathcal{P}_{\mathbb{S}} - \mathcal{P}_{\mathbb{S}} TT^\dagger \mathcal{P}_{\mathbb{S}}\|) \cdot \|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}}^2 \\ &\ge \frac{1}{2}\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}}^2 \text{ (by 6)} \end{aligned} \tag{20}$$

Using the submultiplicative property of operator norm,

$$\|T\mathcal{R}_\Omega T^\dagger - TT^\dagger\| \le \|\mathbf{T}\| \cdot \|\mathcal{R}_\Omega - \mathcal{I}\| \cdot \|\mathbf{T}^\dagger\| \le \frac{\|\mathbf{T}\| \cdot \|\mathbf{T}^\dagger\|}{p}, \tag{21}$$

which gives

$$|B_2| \le \frac{\kappa(\mathbf{T})}{p}\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}}\|\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\|_{\mathrm{F}}. \tag{22}$$

Thus we have

$$\begin{aligned} 0 = |B_1 + B_2| &\ge |B_1| - |B_2| \\ &\ge \frac{1}{2}\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}}^2 - \frac{\|\mathbf{T}\| \cdot \|\mathbf{T}^\dagger\|}{p}\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}}\|\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\|_{\mathrm{F}} \quad (23) \\ &\ge (\frac{1}{2} - 2\frac{p}{8\kappa(\mathbf{T})}\frac{\kappa(\mathbf{T})}{p})\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}}^2 = \frac{1}{4}\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}}^2, \quad (24) \end{aligned}$$

where 24 is followed by case 2 assumption and the inequality between Frobenius norm and nuclear norm that $\|\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\|_{\mathrm{F}} \le \|\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))\|_*$. Thus,

$$\|\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))\|_{\mathrm{F}}^2 \le 0 \Rightarrow \mathcal{P}_{\mathbb{S}}(T(\mathcal{H})) = \mathbf{0}. \tag{25}$$

If $\mathcal{H} \neq \mathbf{0}$, then $T(\mathcal{H}) \neq \mathbf{0}$ due to the injectivity of $T(\cdot)$. Thus $\mathcal{P}_{\mathbb{S}}(T(\mathcal{H}))$ and $\mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H}))$ cannot be $\mathbf{0}$ at the same time following the fact that $\mathcal{P}_{\mathbb{S}}(T(\mathcal{H})) + \mathcal{P}_{\mathbb{S}^\perp}(T(\mathcal{H})) = T(\mathcal{H})$, which further contradicts the assumption of case 2. Thus $\mathcal{H} = \mathbf{0}$.

Combining case 1 and case 2, we obtain that: if $\mathcal{H} \neq \mathbf{0}$, then $\|T(\mathcal{X}+\mathcal{H})\| > \|T(\mathcal{X})\|$, which shows that $\mathcal{X}$ is the unique optimal solution to 1. □

A.4 Construct Dual Certificate

We need to find a construction of dual certificate to prove that under certain conditions, the requirements in Proposition A.1 can be satisfied w.h.p..

Proposition A.3. *If $(\mathcal{X}, T)$ satisfies 2-3 and 4 holds, then 678 and 9 can be satisfied.*

Proof. We show these conditions can be satisfied in turn.

A.5 Show that 6 Can Be Satisfied

We show 6 can be satisfied in steps:

1. Express the operator $\mathcal{P}_{\mathbb{S}} T \mathcal{R}_\Omega T^\dagger \mathcal{P}_{\mathbb{S}}$ as a sum of rank-one operators. Throughout this proof, we use the standard complex Frobenius inner product

$$\langle \mathcal{X}, \mathcal{Y} \rangle := \sum_{i,j,k} \overline{\mathcal{X}_{ijk}} \mathcal{Y}_{ijk}.$$

For any $\mathcal{A}$, by expanding with respect to the canonical basis $\{e_{ijk}\}_{i,j,k}$, we have

$$\begin{aligned} T^\dagger \mathcal{P}_{\mathbb{S}}(\mathcal{A}) &= \sum_{i,j,k} \langle e_{ijk},\ T^\dagger \mathcal{P}_{\mathbb{S}}(\mathcal{A}) \rangle\, e_{ijk} \\ &= \sum_{i,j,k} \langle (T^\dagger)^{\mathrm{H}} e_{ijk},\ \mathcal{P}_{\mathbb{S}}(\mathcal{A}) \rangle\, e_{ijk} \\ &= \sum_{i,j,k} \langle \tilde{T}(e_{ijk}),\ \mathcal{P}_{\mathbb{S}}(\mathcal{A}) \rangle\, e_{ijk} \\ &= \sum_{i,j,k} \langle \mathcal{P}_{\mathbb{S}} \tilde{T}(e_{ijk}),\ \mathcal{A} \rangle\, e_{ijk}, \end{aligned} \tag{26}$$

where $\tilde{T} := (T^\dagger)^{\mathrm{H}}$ and we used the self-adjointness of $\mathcal{P}_{\mathbb{S}}$.

Hence,

$$\mathcal{R}_\Omega T^\dagger \mathcal{P}_{\mathbb{S}}(\mathcal{A}) = \frac{1}{p} \sum_{i,j,k} \delta_{ijk} \langle \mathcal{P}_{\mathbb{S}} \tilde{T}(e_{ijk}),\ \mathcal{A} \rangle\, e_{ijk}, \tag{27}$$

and therefore

$$\mathcal{P}_{\mathbb{S}} T \mathcal{R}_\Omega T^\dagger \mathcal{P}_{\mathbb{S}}(\mathcal{A}) = \frac{1}{p} \sum_{i,j,k} \delta_{ijk} \langle \mathcal{P}_{\mathbb{S}} \tilde{T}(e_{ijk}),\ \mathcal{A} \rangle\, \mathcal{P}_{\mathbb{S}} T(e_{ijk}). \tag{28}$$

Thus,

$$\left(\mathcal{P}_{\mathbb{S}} T \mathcal{R}_\Omega T^\dagger \mathcal{P}_{\mathbb{S}} - \mathcal{P}_{\mathbb{S}} T T^\dagger \mathcal{P}_{\mathbb{S}}\right)(\mathcal{A}) = \sum_{i,j,k} \mathcal{H}_{ijk}(\mathcal{A}), \tag{29}$$

where

$$\mathcal{H}_{ijk} : \mathcal{A} \mapsto \left(\frac{\delta_{ijk}}{p} - 1\right) \langle \mathcal{P}_{\mathbb{S}} \tilde{T}(e_{ijk}),\ \mathcal{A} \rangle\, \mathcal{P}_{\mathbb{S}} T(e_{ijk}). \tag{30}$$

Here $\{\delta_{ijk}\}_{i,j,k}$ are independent Bernoulli random variables with mean p, and $e_{ijk} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is the canonical basis tensor whose (i,j,k)-th entry is 1 and all other entries are 0.

2. Apply the noncommutative Bernstein inequality:

Theorem A.4 (Noncommutative Bernstein inequality (Tropp, 2012))**.** *Let $\{\mathbf{Z}_k\}$ be a finite sequence of independent random matrices of size $n_1 \times n_2$ satisfying $\mathbb{E}[\mathbf{Z}_k] = \mathbf{0}$ and $\|\mathbf{Z}_k\| \le R$ almost surely. Define*

$$\sigma^2 := \max\left\{ \left\| \sum_k \mathbb{E}[\mathbf{Z}_k \mathbf{Z}_k^{\mathrm{H}}] \right\|, \left\| \sum_k \mathbb{E}[\mathbf{Z}_k^{\mathrm{H}} \mathbf{Z}_k] \right\| \right\}. \tag{31}$$

Then for all $t > 0$,

$$\begin{aligned} \mathbb{P}\left[\left\| \sum_k \mathbf{Z}_k \right\| \ge t \right] &\le (n_1 + n_2) \exp\left(-\frac{t^2}{2\sigma^2 + \frac{2}{3} R t} \right) \\ &\le (n_1 + n_2) \exp\left(-\frac{3t^2}{8\sigma^2} \right), \qquad \text{if } t \le \frac{\sigma^2}{R}. \end{aligned}$$

Equivalently, for any $c > 0$, with probability at least $1 - (n_1+n_2)^{1-c}$,

$$\left\|\sum_k \mathbf{Z}_k\right\| \le 2\sqrt{c\sigma^2 \log(n_1+n_2)} + cR\log(n_1+n_2). \tag{32}$$

Now define the block-diagonal matrix operator

$$\mathbf{H}_{ijk}(\cdot) := \mathrm{bdiag}\big(\mathcal{H}_{ijk}(\mathrm{fold}(\cdot))\big).$$

Since $\mathbb{E}[\delta_{ijk}/p - 1] = 0$, we have $\mathbb{E}[\mathcal{H}_{ijk}] = 0$ and hence $\mathbb{E}[\mathbf{H}_{ijk}] = 0$.
For any $\mathcal{A}$,

$$\begin{aligned}\|\mathcal{H}_{ijk}(\mathcal{A})\|_{\mathrm{F}} &= \left\|\left(\frac{\delta_{ijk}}{p} - 1\right)\langle \mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk}),\, \mathcal{A}\rangle\, \mathcal{P}_{\mathbb{S}}T(e_{ijk})\right\|_{\mathrm{F}} \\ &\le \frac{1}{p}\|\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\|_{\mathrm{F}}\|\mathcal{P}_{\mathbb{S}}T(e_{ijk})\|_{\mathrm{F}}\|\mathcal{A}\|_{\mathrm{F}}.\end{aligned} \tag{33}$$

Therefore,

$$\|\mathbf{H}_{ijk}\| \le \frac{1}{p}\|\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\|_{\mathrm{F}}\|\mathcal{P}_{\mathbb{S}}T(e_{ijk})\|_{\mathrm{F}}, \tag{34}$$

and hence

$$R := \max_{i,j,k}\|\mathbf{H}_{ijk}\| \le \frac{1}{p}\max_{i,j,k}\|\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\|_{\mathrm{F}}\max_{i,j,k}\|\mathcal{P}_{\mathbb{S}}T(e_{ijk})\|_{\mathrm{F}}. \tag{35}$$

Next, observe that

$$\mathcal{H}_{ijk}^*\mathcal{H}_{ijk}(\mathcal{A}) = \left(\frac{\delta_{ijk}}{p} - 1\right)^2 \|\mathcal{P}_{\mathbb{S}}T(e_{ijk})\|_{\mathrm{F}}^2\langle \mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk}),\, \mathcal{A}\rangle\, \mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk}). \tag{36}$$

Hence,

$$\begin{aligned}\left\|\sum_{i,j,k}\mathbb{E}[\mathbf{H}_{ijk}^*\mathbf{H}_{ijk}]\right\| &= \left\|\sum_{i,j,k}\mathbb{E}[\mathcal{H}_{ijk}^*\mathcal{H}_{ijk}]\right\| \\ &\le \frac{1-p}{p}\max_{i,j,k}\|\mathcal{P}_{\mathbb{S}}T(e_{ijk})\|_{\mathrm{F}}^2\left\|\sum_{i,j,k}\langle \mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk}),\, \cdot\rangle\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\right\|.\end{aligned} \tag{37}$$

Since

$$\sum_{i,j,k}\langle \mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk}),\, \mathcal{A}\rangle\, \mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk}) = \mathcal{P}_{\mathbb{S}}\tilde{T}\,\tilde{T}^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}(\mathcal{A}), \tag{38}$$

we obtain

$$\left\|\sum_{i,j,k}\mathbb{E}[\mathbf{H}_{ijk}^*\mathbf{H}_{ijk}]\right\| \le \frac{1-p}{p}\max_{i,j,k}\|\mathcal{P}_{\mathbb{S}}T(e_{ijk})\|_{\mathrm{F}}^2\|\mathcal{P}_{\mathbb{S}}\tilde{T}\|^2. \tag{39}$$

Similarly,

$$\left\|\sum_{i,j,k}\mathbb{E}[\mathbf{H}_{ijk}\mathbf{H}_{ijk}^*]\right\| \le \frac{1-p}{p}\max_{i,j,k}\|\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\|_{\mathrm{F}}^2\|\mathcal{P}_{\mathbb{S}}T\|^2. \tag{40}$$

Define

$$\gamma(\mathbf{T}) := \max\left\{\|\mathbf{T}\|^2, \|\mathbf{T}^\dagger\|^2\right\} = \max\left\{\sigma_{\max}^2(\mathbf{T}), \frac{1}{\sigma_{\min}^2(\mathbf{T})}\right\}.$$

Since $\|\mathcal{P}_{\mathbb{S}}T\| \le \|T\|$ and $\|\mathcal{P}_{\mathbb{S}}\tilde{T}\| \le \|\tilde{T}\| = \|T^\dagger\|$, it follows that

$$\sigma^2 \le \frac{1}{p}\,\gamma(\mathbf{T})\,\max\left\{\max_{i,j,k}\|\mathcal{P}_{\mathbb{S}}T(e_{ijk})\|_{\mathrm{F}}^2,\, \max_{i,j,k}\|\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\|_{\mathrm{F}}^2\right\}. \tag{41}$$

Applying Theorem A.4, we have

$$\begin{aligned}&\mathbb{P}\left[\left\|\mathcal{P}_{\mathbb{S}}T\mathcal{R}_{\Omega}T^{\dagger}\mathcal{P}_{\mathbb{S}}-\mathcal{P}_{\mathbb{S}}TT^{\dagger}\mathcal{P}_{\mathbb{S}}\right\|\geq t\right]\\&=\mathbb{P}\left[\left\|\sum_{i,j,k}\mathbf{H}_{ijk}\right\|\geq t\right]\\&\leq(n_1+n_2)N_3\exp\left(-\frac{t^2}{2\sigma^2+\frac{2}{3}Rt}\right).\end{aligned}\tag{42}$$

Let $t=\frac{1}{2}$. By Appendix C,

$$\|\mathcal{P}_{\mathbb{S}}T(e_{ijk})\|_{\mathrm{F}}^2\leq\frac{\nu r(n_1+n_2)}{n_1n_2}\|\mathbf{T}\|_{1\to2}^2,\qquad\|\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\|_{\mathrm{F}}^2\leq\frac{\nu r(n_1+n_2)}{n_1n_2}\|\tilde{\mathbf{T}}\|_{1\to2}^2.$$

Using $\|\mathbf{A}\|_{\infty,2}\leq\|\mathbf{A}\|$ for any matrix $\mathbf{A}$, and the sampling rate condition in 4, there exist constants $c_3>0$ and $\beta>1$ such that

$$\mathbb{P}\left[\left\|\mathcal{P}_{\mathbb{S}}T\mathcal{R}_{\Omega}T^{\dagger}\mathcal{P}_{\mathbb{S}}-\mathcal{P}_{\mathbb{S}}TT^{\dagger}\mathcal{P}_{\mathbb{S}}\right\|\geq\frac{1}{2}\right]\leq c_3^{-\beta}\big((n_1+n_2)N_3\big)^{1-\beta},\tag{43}$$

which implies that 6 holds with high probability.

A.6 Prove that 7, 8, 9 Can Be Satisfied

Use Golfing Scheme (Gross, 2011) and construct $\mathcal{Y}$ as

$$\mathcal{Y}=\sum_{i=1}^{l}(\frac{1}{p_i}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}}+\mathcal{I}-\tilde{T}T^{\mathrm{H}})\mathcal{M}_{i-1},\tag{44}$$

$$\mathcal{M}_i=\mathcal{P}_{\mathbb{S}}(\tilde{T}T^{\mathrm{H}}-\frac{1}{p_i}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}})\mathcal{M}_{i-1},\tag{45}$$

$$\mathcal{M}_0=\mathcal{U}\star\mathcal{V}^{\mathrm{H}},\tag{46}$$

and $\{\Omega_i\}_{i=1}^l$ are l independent sets of indices sampled by $\mathrm{Ber}(p_i), p_i=1-(1-p)^{\frac{1}{l}}$, $\bigcup_{i=1}^l\Omega_i=\Omega$, $l=10\log(\gamma(\mathbf{T})(n_1+n_2)N_3)$.

A.6.1 Prove that 7 Can Be Satisfied

First we need to prove the construction 44, 45, 46 satisfies 7:

$$\begin{aligned}&(TT^{\dagger}-T\mathcal{S}_{\Omega}T^{\dagger})^{\mathrm{H}}(\frac{1}{p_i}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}}+\mathcal{I}-\tilde{T}T^{\mathrm{H}})\\=&\frac{1}{p_i}(TT^{\dagger}-T\mathcal{S}_{\Omega}T^{\dagger})^{\mathrm{H}}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}}+(TT^{\dagger}-T\mathcal{S}_{\Omega}T^{\dagger})^{\mathrm{H}}(\mathcal{I}-\tilde{T}T^{\mathrm{H}})\\=&\mathbf{0}\ (\text{by }\mathbf{T}^{\mathrm{H}}\tilde{\mathbf{T}}=(\mathbf{T}^{\dagger}\mathbf{T})^{\mathrm{H}}=\mathbf{I},\mathcal{S}_{\Omega}\mathcal{S}_{\Omega_i}=\mathcal{S}_{\Omega_i}),\end{aligned}\tag{47}$$

thus 7 is satisfied.

A.6.2 Prove that 8 Can Be Satisfied

Following the construction,

$$\begin{aligned}\mathcal{P}_{\mathbb{S}}(\mathcal{Y})&=\sum_{i=1}^{l}\mathcal{P}_{\mathbb{S}}(\frac{1}{p_i}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}}+\mathcal{I}-\tilde{T}T^{\mathrm{H}})(\mathcal{M}_{i-1})\\&=\sum_{i=1}^{l}\mathcal{P}_{\mathbb{S}}\mathcal{M}_{i-1}-\mathcal{P}_{\mathbb{S}}(\tilde{T}T^{\mathrm{H}}-\frac{1}{p_i}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}})\mathcal{M}_{i-1}\\&=\sum_{i=1}^{l}\mathcal{M}_{i-1}-\mathcal{M}_i=\mathcal{U}\star\mathcal{V}^{\mathrm{H}}-\mathcal{M}_l.\end{aligned}\tag{48}$$

Therefore, $\|\mathcal{P}_{\mathbb{S}}(\mathcal{Y}) - \mathcal{U} \star \mathcal{V}^{\mathrm{H}}\|_{\mathrm{F}} = \|\mathcal{M}_l\|_{\mathrm{F}}$.

Since $\mathcal{M}_0 \in \mathbb{S}$, we desire that

$$\begin{aligned} \|\mathcal{M}_i\|_{\mathrm{F}} &= \|\mathcal{P}_{\mathbb{S}}(\tilde{T}\,T^{\mathrm{H}} - \frac{1}{p_i}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}})\mathcal{M}_{i-1}\|_{\mathrm{F}} \\ &\leq \|\mathcal{P}_{\mathbb{S}}\tilde{T}\,T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}} - \frac{1}{p_i}\mathcal{P}_{\mathbb{S}}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\| \cdot \|\mathcal{M}_{i-1}\|_{\mathrm{F}} \leq \frac{1}{2}\|\mathcal{M}_{i-1}\|_{\mathrm{F}}. \end{aligned} \tag{49}$$

Note that 49 puts a requirement similar to 6 on the per-step sampling rate p_i. Also, $p_i = 1-(1-p)^{\frac{1}{l}} \geq \frac{p}{l}, l = 10\log(\gamma(\mathbf{T})(n_1+n_2)N_3)$, thus $\exists\ c_4 > 0$,

$$p_i \geq c_4(\gamma^2(\mathbf{T}) + \rho^2(\mathbf{T}))\frac{\lambda r(n_1+n_2)}{n_1 n_2}\log(\gamma(\mathbf{T})(n_1+n_2)N_3). \tag{50}$$

Using 42, $\exists\ c_5 > 0, \beta^{'} > 1$ by union bound $\|\mathcal{M}_i\|_{\mathrm{F}} \leq \frac{1}{2}\|\mathcal{M}_{i-1}\|_{\mathrm{F}}$ holds for all $i = 1, 2, \ldots, l$ with probability at least $1 - c_5((n_1+n_2)N_3)^{1-\beta^{'}}$, which ensures 49 holds.

As $l = 10\log(\gamma(\mathbf{T})(n_1+n_2)N_3)$, then w.h.p.

$$\|\mathcal{P}_{\mathbb{S}}(\mathcal{Y}) - \mathcal{U} \star \mathcal{V}^{\mathrm{H}}\|_{\mathrm{F}} = \|\mathcal{M}_l\|_{\mathrm{F}} \leq (\frac{1}{2})^l\|\mathcal{U} \star \mathcal{V}\|_{\mathrm{F}} \leq (\frac{1}{2})^l\sqrt{rN_3} \leq \frac{p}{8\kappa(\mathbf{T})}. \tag{51}$$

A.6.3 Prove that 9 Can Be Satisfied

Now we prove 9.

First we use the construction 44 and the triangle inequality:

$$\begin{aligned} \|\mathcal{P}_{\mathbb{S}^{\perp}}(\mathcal{Y})\| &= \|\sum_{i=1}^{l}\mathcal{P}_{\mathbb{S}^{\perp}}(\frac{1}{p_i}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}} - \tilde{T}\,T^{\mathrm{H}})(\mathcal{M}_{i-1})\| \\ &\leq \sum_{i=1}^{l}\|\mathcal{P}_{\mathbb{S}^{\perp}}(\frac{1}{p_i}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}} - \tilde{T}\,T^{\mathrm{H}})(\mathcal{M}_{i-1})\| \\ &\leq \sum_{i=1}^{l}\|(\frac{1}{p_i}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}} - \tilde{T}\,T^{\mathrm{H}})(\mathcal{M}_{i-1})\|. \end{aligned} \tag{52}$$

Here we need a lemma for the initial control of the spectral norm and the lemma is proved in Appendix B.1.

Lemma A.5. *Assuming that the Bernoulli sampling scheme is employed and* $\forall c > 0$,

$$\|\tilde{T}(\frac{1}{p}\mathcal{S}_{\Omega} - \mathcal{I})\mathcal{Z}\| \leq 2\sqrt{\frac{c}{p}\log((n_1+n_2)N_3)}\|\tilde{\mathbf{T}}\|_{\infty}\|\mathcal{Z}\|_{\infty,2} + \frac{c}{p}\log((n_1+n_2)N_3)\|\tilde{\mathbf{T}}\|_{\infty}\|\mathcal{Z}\|_{\infty}$$

holds with probability at least $1-(n_1+n_2)^{1-c}$.

By Lemma A.5, for $c_5 > 1$, w.h.p.

$$\begin{aligned} \|\mathcal{P}_{\mathbb{S}^{\perp}}(\mathcal{Y})\| &\leq \sum_{i=1}^{l}\|(\frac{1}{p_i}\tilde{T}\mathcal{S}_{\Omega_i}T^{\mathrm{H}} - \tilde{T}\,T^{\mathrm{H}})(\mathcal{M}_{i-1})\| \\ &\leq c_5\sum_{i=1}^{l}\frac{1}{p_i}\log((n_1+n_2)N_3)\|\tilde{\mathbf{T}}\|_{\infty}\|\,T^{\mathrm{H}}(\mathcal{M}_{i-1})\|_{\infty} \\ &+2\sqrt{\frac{c_5}{p_i}\log((n_1+n_2)N_3)}\|\tilde{\mathbf{T}}\|_{\infty}\|\,T^{\mathrm{H}}(\mathcal{M}_{i-1})\|_{\infty,2}. \end{aligned} \tag{53}$$

Now another lemma is introduced to control the $\|\cdot\|_{\infty}$ norm and we prove it in Appendix B.2.

Lemma A.6. *Assuming that the Bernoulli sampling scheme is employed, if* $p \geq c_0\gamma^2(\mathbf{T})\frac{\nu r(n_1+n_2)}{n_1n_2}\log((n_1+n_2)N_3)$*, then*

$$\|T^H\mathcal{P}_{\mathbb{S}}\tilde{T}(\mathcal{I}-\frac{1}{p}\mathcal{S}_\Omega)\mathcal{Z}\|_\infty \leq \frac{1}{2}\|\mathcal{Z}\|_\infty$$

holds with probability at least $1-2(n_1+n_2)^{2-\frac{3c_0}{32}}N_3^{1-\frac{3c_0}{32}}$.

By Lemma A.6 and 50, w.h.p.

$$\begin{aligned}\|T^{\mathrm{H}}(\mathcal{M}_{i-1})\|_\infty &= \|T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(\mathcal{I}-\frac{1}{p_i}\mathcal{S}_{\Omega_i})T^{\mathrm{H}}\mathcal{M}_{i-2}\|_\infty \\ &\leq \frac{1}{2}\|T^{\mathrm{H}}(\mathcal{M}_{i-2})\|_\infty \leq (\frac{1}{2})^{i-1}\|T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_\infty.\end{aligned} \tag{54}$$

Also a lemma that controls $\|\cdot\|_{\infty,2}$ norm is needed and the proof is given in Appendix B.3.

Lemma A.7. *Assuming that the Bernoulli sampling scheme is employed, if* $p \geq c_0(\gamma^2(\mathbf{T})+\rho^2(\mathbf{T}))\frac{\lambda r(n_1+n_2)}{n_1n_2}\log((n_1+n_2)N_3)$ *with* c_0 *large enough, then for some* $c>1$*,*

$$\|T^H\mathcal{P}_{\mathbb{S}}\tilde{T}(\mathcal{I}-\frac{1}{p}\mathcal{S}_\Omega)\mathcal{Z}\|_{\infty,2} \leq \frac{1}{2}\|\mathcal{Z}\|_{\infty,2}+\frac{1}{2}\sqrt{\frac{n_1n_2}{\mu r(n_1+n_2)}}\|\mathcal{Z}\|_\infty$$

holds with probability at least $1-2((n_1+n_2)n_3)^{1-c}$.

By Lemma A.7, w.h.p.

$$\begin{aligned}&\|T^{\mathrm{H}}(\mathcal{M}_{i-1})\|_{\infty,2}\\ =&\|T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(\mathcal{I}-\frac{1}{p_i}\mathcal{S}_{\Omega_i})T^{\mathrm{H}}(\mathcal{M}_{i-2})\|_{\infty,2}\\ \leq&\frac{1}{2}\sqrt{\frac{n_1n_2}{\mu r(n_1+n_2)}}\|T^{\mathrm{H}}(\mathcal{M}_{i-2})\|_\infty+\frac{1}{2}\|T^{\mathrm{H}}(\mathcal{M}_{i-2})\|_{\infty,2}\\ \leq&(\frac{1}{2})^2\sqrt{\frac{n_1n_2}{\mu r(n_1+n_2)}}\|T^{\mathrm{H}}(\mathcal{M}_{i-3})\|_\infty\\ &+\frac{1}{2}(\frac{1}{2}\sqrt{\frac{n_1n_2}{\mu r(n_1+n_2)}}\|T^{\mathrm{H}}(\mathcal{M}_{i-3})\|_\infty+\frac{1}{2}\|T^{\mathrm{H}}(\mathcal{M}_{i-3})\|_{\infty,2})\\ \leq&(i-1)(\frac{1}{2})^{i-1}\sqrt{\frac{n_1n_2}{\mu r(n_1+n_2)}}\|T^{\mathrm{H}}\mathcal{M}_0\|_\infty+(\frac{1}{2})^{i-1}\|T^{\mathrm{H}}\mathcal{M}_0\|_{\infty,2}\end{aligned} \tag{55}$$

Plug 54 and 55 into 53, yielding

$$\begin{aligned}\|\mathcal{P}_{\mathbb{S}^\perp}(\mathcal{Y})\| \leq c_5\|\tilde{\mathbf{T}}\|_\infty\frac{\log((n_1+n_2)N_3)}{p_i}\sum_{i=1}^{l}\Big(\frac{1}{2}\Big)^{i-1}\|T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_\infty\\ +2\sqrt{c_5}\,\|\tilde{\mathbf{T}}\|_\infty\sqrt{\frac{\log((n_1+n_2)N_3)}{p_i}}\sum_{i=1}^{l}\Bigg[(i-1)\Big(\frac{1}{2}\Big)^{i-1}\sqrt{\frac{n_1n_2}{\mu r(n_1+n_2)}}\|T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_\infty\\ +\Big(\frac{1}{2}\Big)^{i-1}\|T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_{\infty,2}\Bigg].\end{aligned}$$

Using

$$\sum_{i=1}^{l}\Big(\frac{1}{2}\Big)^{i-1}\leq 2, \qquad \sum_{i=1}^{l}(i-1)\Big(\frac{1}{2}\Big)^{i-1}\leq 2,$$

we further obtain

$$\begin{aligned}\|\mathcal{P}_{\mathbb{S}^\perp}(\mathcal{Y})\| \le 2c_5\|\tilde{\mathbf{T}}\|_\infty \frac{\log((n_1+n_2)N_3)}{p_i}\|\,T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_\infty \\ + 4\sqrt{c_5}\,\|\tilde{\mathbf{T}}\|_\infty\sqrt{\frac{n_1 n_2\log((n_1+n_2)N_3)}{\mu r(n_1+n_2)p_i}}\|\,T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_\infty \\ + 4\sqrt{c_5}\,\|\tilde{\mathbf{T}}\|_\infty\sqrt{\frac{\log((n_1+n_2)N_3)}{p_i}}\|\,T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_{\infty,2}.\end{aligned} \tag{56}$$

Now it remains to bound $\|\,T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_\infty$ and $\|\,T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_{\infty,2}$.

Here, in order to bound $\|\,T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_\infty$, a domain transform strategy is employed and we revisit the definition of generalized convolution $\star_{\mathbf{A},\mathbf{B}}$ induced by two transforms $\mathbf{A},\mathbf{B}$ (Korohoda and Dąbrowski, 2004), which gives the form induced by $T(\cdot)$ of each entry in the original domain:

Theorem A.8 (Generalized convolution)**.** *(Korohoda and Dąbrowski, 2004) Suppose* $\mathbf{x},\mathbf{y}\in\mathbb{C}^{N\times 1}$, $\mathbf{A}\in\mathbb{C}^{M\times N}$, *and* $\mathbf{B}\in\mathbb{C}^{L\times M}$. *Define*

$$\mathbf{z}=\mathbf{B}\big((\mathbf{A}\mathbf{x})\star(\mathbf{A}\mathbf{y})\big),$$

where $\star$ *denotes the entrywise product. Then for each* $\ell=1,\dots,L$,

$$\mathbf{z}(\ell)=\mathbf{x}^{\mathrm{T}}\mathbf{W}^{\ell}_{\mathbf{A},\mathbf{B}}\mathbf{y},$$

where $\mathbf{W}^{\ell}_{\mathbf{A},\mathbf{B}}\in\mathbb{C}^{N\times N}$ *is given by*

$$\mathbf{W}^{\ell}_{\mathbf{A},\mathbf{B}}(i,j)=\sum_{k=1}^{M}\mathbf{B}(\ell,k)\mathbf{A}(k,i)\mathbf{A}(k,j). \tag{57}$$

Let $\mathbf{u}_t^i=\mathcal{U}(i,t,:)$ and $(\mathbf{v}_t^j)^{\mathrm{H}}=\mathcal{V}^{\mathrm{H}}(t,j,:)$. Then with Theorem A.8,

$$T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})(i,j,:)=\sum_{t=1}^{r}\mathbf{u}_t^i\star_{\mathbf{I},\mathbf{T}^{\mathrm{H}}}(\mathbf{v}_t^j)^{\mathrm{H}}.$$

Hence,

$$\begin{aligned}\|\,T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_\infty &\le \max_{i,j}\sum_{t=1}^{r}\|\mathbf{u}_t^i\star_{\mathbf{I},\mathbf{T}^{\mathrm{H}}}(\mathbf{v}_t^j)^{\mathrm{H}}\|_\infty \\ &\le \max_{i,j}\sum_{t=1}^{r}\max_k\|W_T^k\|\,\|\mathbf{u}_t^i\|_2\,\|\mathbf{v}_t^j\|_2 \\ &\le \max_k\|W_T^k\|\max_{i,j}\sum_{t=1}^{r}\left(\frac{1}{2}\|\mathbf{u}_t^i\|_2^2+\frac{1}{2}\|\mathbf{v}_t^j\|_2^2\right) \\ &= \max_k\|W_T^k\|\max_{i,j}\left\{\frac{1}{2}\|\mathcal{U}^{\mathrm{H}}\star\xi_i\|_F^2+\frac{1}{2}\|\mathcal{V}^{\mathrm{H}}\star\xi_j\|_F^2\right\}.\end{aligned}$$

Since $W_{\mathbf{T}}^k=\mathrm{diag}(T^{\mathrm{H}}(k,:))$, we have

$$\max_k\|W_{\mathbf{T}}^k\|=\|\mathbf{T}^{\mathrm{H}}\|_\infty=\|\mathbf{T}\|_\infty.$$

Therefore,

$$\|\,T^{\mathrm{H}}(\mathcal{U}\star\mathcal{V}^{\mathrm{H}})\|_\infty\le\|\mathbf{T}\|_\infty\frac{\mu r(n_1+n_2)N_3}{2n_1n_2}. \tag{58}$$

Another reason why we introduce Theorem A.8 here is that when $\mathbf{T}$ is a square matrix (not necessarily orthogonal), by letting $\mathbf{B} = \tilde{\mathbf{T}}$ the bound can be traced back to the singular tensors in the original domain and possibly some tighter bound can be obtained.

For $\| T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\infty,2}$,

$$\begin{aligned}\| T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\infty,2} &\le \max_{ij}\{\| T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}}) \star e_j\|_{\mathrm{F}}, \|e_i^{\mathrm{H}} \star T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\mathrm{F}}\} \\ &= \max_{ij}\{\| T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}} \star \xi_j)\|_{\mathrm{F}}, \| T^{\mathrm{H}}(\xi_i^{\mathrm{H}} \star \mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\mathrm{F}}\} \\ &\le \|\mathbf{T}^{\mathrm{H}}\|\sqrt{\frac{\mu r(n_1+n_2)N_3}{n_1 n_2}}.\end{aligned} \tag{59}$$

Thus, plug $p_i \ge \frac{c_0}{10}(\gamma^2(\mathbf{T}) + \rho^2(\mathbf{T}))\frac{\lambda r(n_1+n_2)}{n_1 n_2}\log((n_1+n_2)N_3)$ and 58, 59 into 56 and we finally obtain w.h.p.

$$\begin{aligned}&\|\mathcal{P}_{\mathbb{S}^\perp}(\mathcal{Y})\| \\ \le&\frac{c_5}{p_i}\cdot\frac{\rho(\mathbf{T})\mu r(n_1+n_2)\log((n_1+n_2)N_3)}{2n_1n_2} + 8\sqrt{c_5}\rho(\mathbf{T})\sqrt{\frac{\mu r(n_1+n_2)\log((n_1+n_2)N_3)}{2n_1n_2p_i}} \\ &+ 4\sqrt{c_5}\sqrt{\frac{(\kappa^2(\mathbf{T})+\rho^2(\mathbf{T}))\mu r(n_1+n_2)}{n_1n_2p_i}} \le \frac{10c_5}{c_0} + \frac{32\sqrt{c_5}}{\sqrt{c_0}} + \frac{16\sqrt{c_5}}{\sqrt{c_0}} \le \frac{1}{2}\end{aligned} \tag{60}$$

with c_0 large enough. □

B Proofs of the Norm-bounding Lemmas

B.1 Proof of Lemma A.5

Proof. First, we decompose the operator,

$$\tilde{T}(\frac{1}{p}\mathcal{S}_\Omega - \mathcal{I})\mathcal{Z} = \sum_{ijk}(\frac{\delta_{ijk}}{p} - 1)\mathcal{Z}_{ijk}\tilde{T}(e_{ijk}) := \sum_{ijk}\mathcal{L}_{ijk}. \tag{61}$$

With $\mathbb{E}\left[\mathtt{bdiag}(\mathcal{L}_{ijk})\right] = \mathbf{0}$,

$$\|\mathtt{bdiag}(\mathcal{L}_{ijk})\| \le \frac{1}{p}\|\mathcal{Z}\|_\infty \max_{ijk}\|\mathtt{bdiag}(\tilde{T}(e_{ijk}))\| \le \frac{1}{p}\|\mathcal{Z}\|_\infty\|\tilde{\mathbf{T}}\|_\infty \tag{62}$$

$$\begin{aligned}&\|\sum_{ijk}\mathbb{E}[\mathtt{bdiag}^{\mathrm{H}}(\mathcal{L}_{ijk})\mathtt{bdiag}(\mathcal{L}_{ijk})]\| \\ =&\|\sum_{ijk}\mathbb{E}\left[(\frac{\delta_{ijk}}{p}-1)^2\right]\cdot|\mathcal{Z}_{ijk}|^2\cdot\mathtt{bdiag}(\left[\tilde{T}(e_{ijk})\right]^{\mathrm{H}}\star\tilde{T}(e_{ijk}))\| \\ =&\frac{1-p}{p}\|\sum_{ijk}|\mathcal{Z}_{ijk}|^2\cdot\left[\tilde{T}(e_{ijk})\right]^{\mathrm{H}}\tilde{T}(e_{ijk})\| \\ \le&\frac{1-p}{p}\|\tilde{\mathbf{T}}\|_\infty^2\max_j\sum_{ik}|\mathcal{Z}_{ijk}|^2 \\ =&\frac{1-p}{p}\|\tilde{\mathbf{T}}\|_\infty^2\|\mathcal{Z}_{ijk}\|_{1\mapsto 2}^2\end{aligned} \tag{63}$$

$$\|\sum_{ijk}\mathbb{E}[\mathtt{bdiag}(\mathcal{L}_{ijk})\mathtt{bdiag}^{\mathrm{H}}(\mathcal{L}_{ijk})]\| \le \frac{1-p}{p}\|\tilde{\mathbf{T}}\|_\infty^2\|\mathcal{Z}_{ijk}\|_{2\mapsto\infty}^2 \tag{64}$$

Use the extension of theorem A.4:

$$\begin{aligned}\|\tilde{T}(\frac{1}{p}\mathcal{S}_\Omega - \mathcal{I})\mathcal{Z}\| &= \|\sum_{ijk}\mathcal{L}_{ijk}\| = \|\sum_{ijk}\texttt{bdiag}(\mathcal{L}_{ijk})\| \\ &\leq 2\sqrt{c\frac{1}{p}\log((n_1+n_2)N_3)}\|\tilde{\mathbf{T}}\|_\infty\|\mathcal{Z}\|_{\infty,2} + c\frac{1}{p}\log((n_1+n_2)N_3)\|\tilde{\mathbf{T}}\|_\infty\|\mathcal{Z}\|_\infty\end{aligned} \tag{65}$$

holds with probability at least $1-(n_1+n_2)^{1-c}$, $\forall c>0$. □

B.2 Proof of Lemma A.6

Proof. First we decompose the operator:

$$\begin{aligned}T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}\Big(\mathcal{I}-\frac{1}{p}\mathcal{S}_\Omega\Big)\mathcal{Z} &= T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}\sum_{i,j,k}\Big(1-\frac{\delta_{ijk}}{p}\Big)\mathcal{Z}_{ijk}e_{ijk} \\ &= \sum_{i,j,k}\Big(1-\frac{\delta_{ijk}}{p}\Big)\mathcal{Z}_{ijk}\ T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk}) =: \sum_{i,j,k}\mathcal{A}_{ijk}.\end{aligned} \tag{66}$$

Each entry of $\mathcal{A}_{ijk}$ can be extracted via the inner product with e_{abc}:

$$\begin{aligned}\mathcal{A}_{ijk,abc} &= \Big\langle e_{abc}, \Big(1-\frac{\delta_{ijk}}{p}\Big)\mathcal{Z}_{ijk}\ T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\Big\rangle \\ &= \Big(1-\frac{\delta_{ijk}}{p}\Big)\mathcal{Z}_{ijk}\left\langle\mathcal{P}_{\mathbb{S}}T(e_{abc}), \mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\right\rangle.\end{aligned} \tag{67}$$

Therefore,

$$\begin{aligned}|\mathcal{A}_{ijk,abc}| &= \left|\Big(1-\frac{\delta_{ijk}}{p}\Big)\mathcal{Z}_{ijk}\left\langle\mathcal{P}_{\mathbb{S}}T(e_{abc}), \mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\right\rangle\right| \\ &\leq \frac{1}{p}\|\mathcal{Z}\|_\infty\|\mathcal{P}_{\mathbb{S}}T(e_{abc})\|_{\mathrm{F}}\|\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\|_{\mathrm{F}}.\end{aligned} \tag{68}$$

Moreover,

$$\begin{aligned}\sum_{i,j,k}\mathbb{E}\big[|\mathcal{A}_{ijk,abc}|^2\big] &= \sum_{i,j,k}\mathbb{E}\left[\Big(1-\frac{\delta_{ijk}}{p}\Big)^2|\mathcal{Z}_{ijk}|^2\left|\left\langle\mathcal{P}_{\mathbb{S}}T(e_{abc}), \mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\right\rangle\right|^2\right] \\ &= \sum_{i,j,k}\frac{1-p}{p}|\mathcal{Z}_{ijk}|^2\left|\left\langle\mathcal{P}_{\mathbb{S}}T(e_{abc}), \mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\right\rangle\right|^2 \\ &\leq \frac{1-p}{p}\|\mathcal{Z}\|_\infty^2\sum_{i,j,k}\left|\left\langle\mathcal{P}_{\mathbb{S}}T(e_{abc}), \mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\right\rangle\right|^2 \\ &= \frac{1-p}{p}\|\mathcal{Z}\|_\infty^2\|\,T^\dagger\mathcal{P}_{\mathbb{S}}T(e_{abc})\|_{\mathrm{F}}^2 \\ &\leq \frac{1-p}{p}\|\mathcal{Z}\|_\infty^2\|\mathbf{T}^\dagger\|^2\|\mathcal{P}_{\mathbb{S}}T(e_{abc})\|_{\mathrm{F}}^2.\end{aligned} \tag{69}$$

Define

$$\gamma(\mathbf{T}) := \max\left\{\|\mathbf{T}\|^2, \|\mathbf{T}^\dagger\|^2\right\} = \max\left\{\sigma_{\max}^2(\mathbf{T}), \frac{1}{\sigma_{\min}^2(\mathbf{T})}\right\}.$$

Then, by Appendix C,

$$\|\,T^\dagger\mathcal{P}_{\mathbb{S}}T(e_{abc})\|_{\mathrm{F}}^2 \leq \gamma(\mathbf{T})\,\frac{\nu r(n_1+n_2)}{n_1n_2}.$$

Using Bernstein's inequality,

$$
\begin{aligned}
&\mathbb{P}\left[\left|\left(T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(\mathcal{I}-\tfrac{1}{p}\mathcal{S}_{\Omega})\mathcal{Z}\right)_{abc}\right| \geq \frac{1}{2}\|\mathcal{Z}\|_{\infty}\right] \\
&= \mathbb{P}\left[\left|\sum_{i,j,k}\mathcal{A}_{ijk,abc}\right| \geq \frac{1}{2}\|\mathcal{Z}\|_{\infty}\right] \\
&\leq 2\exp\left(-\frac{\frac{1}{4}\|\mathcal{Z}\|_{\infty}^2}{2\sum_{i,j,k}\mathbb{E}[|\mathcal{A}_{ijk,abc}|^2]+\frac{2}{3}\cdot\frac{1}{2}\|\mathcal{Z}\|_{\infty}\cdot\max_{i,j,k}|\mathcal{A}_{ijk,abc}|}\right) \\
&\leq 2\exp\left(-\frac{\frac{1}{4}\|\mathcal{Z}\|_{\infty}^2}{\frac{2}{p}\|\mathcal{Z}\|_{\infty}^2\|T^{\dagger}\mathcal{P}_{\mathbb{S}}T(e_{abc})\|_{\mathrm{F}}^2+\frac{1}{3p}\|\mathcal{Z}\|_{\infty}^2\|\mathcal{P}_{\mathbb{S}}T(e_{abc})\|_{\mathrm{F}}\max_{i,j,k}\|\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\|_{\mathrm{F}}}\right) \\
&= 2\exp\left(-\frac{p}{8\|T^{\dagger}\mathcal{P}_{\mathbb{S}}T(e_{abc})\|_{\mathrm{F}}^2+\frac{4}{3}\|\mathcal{P}_{\mathbb{S}}T(e_{abc})\|_{\mathrm{F}}\max_{i,j,k}\|\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\|_{\mathrm{F}}}\right) \\
&\leq 2\exp\left(-\frac{c_0\gamma(\mathbf{T})\frac{\nu r(n_1+n_2)}{n_1n_2}\log((n_1+n_2)N_3)}{8\gamma(\mathbf{T})\frac{\nu r(n_1+n_2)}{n_1n_2}+\frac{4}{3}\gamma(\mathbf{T})\frac{\nu r(n_1+n_2)}{n_1n_2}}\right) \\
&\leq 2((n_1+n_2)N_3)^{-\frac{3c_0}{28}}.
\end{aligned}
\tag{70}
$$

By the union bound,

$$
\|T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(\mathcal{I}-\tfrac{1}{p}\mathcal{S}_{\Omega})\mathcal{Z}\|_{\infty} \leq \|\mathcal{Z}\|_{\infty}
$$

holds with probability at least

$$
1-2(n_1+n_2)^{1-\frac{3c_0}{28}}N_3^{1-\frac{3c_0}{28}}.
$$

□

B.3 Proof of Lemma A.7

Proof. We decompose the operator as

$$
\begin{aligned}
&(T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(\mathcal{I}-\frac{1}{p}\mathcal{S}_{\Omega})\mathcal{Z})\star\bar{\xi}_b \\
=&\sum_{ijk}(1-\frac{\delta_{ijk}}{p})\mathcal{Z}_{ijk}(T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk}))\star\bar{\xi}_b := \sum_{ijk}\mathcal{C}_{ijk}\in\mathbb{C}^{n_1\times 1\times n_3},
\end{aligned}
\tag{71}
$$

where $\bar{\xi}_b \in \mathbb{C}^{n_2\times 1\times n_3}$ is the tensor column basis in the original domain with $\bar{\xi}_b(b,1,:)=1$ and equaling 0 otherwise. Since $\|\mathcal{C}_{ijk}\|_{\mathrm{F}} = \|\mathtt{vec}(\mathcal{C}_{ijk})\|_2 = \|\mathtt{vec}(\mathcal{C}_{ijk})\|$, it is equipped with the form to apply Theorem A.4.

$$
\begin{aligned}
&\|\mathtt{vec}(\mathcal{C}_{ijk})\|_2 = \|\mathcal{C}_{ijk}\|_{\mathrm{F}} \\
&\leq \frac{1}{p}\|\mathcal{Z}\|_{\infty}\|(T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk}))\star\bar{\xi}_b\|_{\mathrm{F}} \leq \frac{1}{p}\|\mathcal{Z}\|_{\infty}\|T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk})\|_{\mathrm{F}}.
\end{aligned}
\tag{72}
$$

$$
\begin{aligned}
&|\mathbb{E}[\sum_{ijk}\mathtt{vec}(\mathcal{C}_{ijk})^{\mathrm{H}}\mathtt{vec}(\mathcal{C}_{ijk})]| = |\mathbb{E}[\sum_{ijk}\|\mathcal{C}_{ijk}\|_{\mathrm{F}}^2]| \\
=&\frac{1-p}{p}\sum_{ijk}|\mathcal{Z}_{ijk}|^2\|(T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk}))\star\bar{\xi}_b\|_{\mathrm{F}}^2
\end{aligned}
\tag{73}
$$

Looking into the term in 73:

$$
\|(T^{\mathrm{H}}\mathcal{P}_{\mathbb{S}}\tilde{T}(e_{ijk}))\star\bar{\xi}_b)\|_{\mathrm{F}} \leq \|(T^{\mathrm{H}}(\mathcal{P}_{\mathcal{U}}\star\tilde{T}(e_{ijk})))\star\bar{\xi}_b\|_{\mathrm{F}} + \|(T^{\mathrm{H}}(\mathcal{P}_{\mathcal{U}^{\perp}}\star\tilde{T}(e_{ijk})\star\mathcal{P}_{\mathcal{V}})\star\bar{\xi}_b\|_{\mathrm{F}},
\tag{74}
$$

yields

$$\|(T^{\mathrm{H}}(\mathcal{P}_{\mathcal{U}} \star \tilde{T}(e_{ijk}))) \star \bar{\xi}_b\|_{\mathrm{F}} = \begin{cases} \|T^{\mathrm{H}}(\mathcal{P}_{\mathcal{U}} \star \tilde{T}(e_{ibk}))\|_{\mathrm{F}} & , \ j = b \\ 0 & , \ j \neq b \end{cases} \tag{75}$$

$$\|(T^{\mathrm{H}}(\mathcal{P}_{\mathcal{U}^\perp} \star \tilde{T}(e_{ijk}) \star \mathcal{P}_{\mathcal{V}}) \star \bar{\xi}_b\|_{\mathrm{F}} \leq \|\mathbf{T}^{\mathrm{H}}\| \cdot \|\tilde{\mathbf{T}}\|_\infty \|\xi_j^{\mathrm{H}} \star \mathcal{P}_{\mathcal{V}} \star \xi_b\|_{\mathrm{F}}. \tag{76}$$

Therefore

$$\begin{aligned}
&|\mathbb{E}[\sum_{ijk} \mathtt{vec}(\mathcal{C}_{ijk})^{\mathrm{H}} \mathtt{vec}(\mathcal{C}_{ijk})]| = |\mathbb{E}[\sum_{ijk} \|\mathcal{C}_{ijk}\|_{\mathrm{F}}^2]| \\
\leq &\frac{2}{p} \sum_{ik} |\mathcal{Z}_{ibk}|^2 \|T^{\mathrm{H}}(\mathcal{P}_{\mathcal{U}} \star \tilde{T}(e_{ibk}))\|_{\mathrm{F}}^2 + \frac{2}{p} \|\mathbf{T}^{\mathrm{H}}\|^2 \cdot \|\tilde{\mathbf{T}}\|_\infty^2 \sum_{ijk} |\mathcal{Z}_{ijk}|^2 \cdot \|\xi_j^{\mathrm{H}} \star \mathcal{P}_{\mathcal{V}} \star \xi_b\|_{\mathrm{F}}^2 \\
\leq &\frac{2}{p} \max_{ijk} \|T^{\mathrm{H}} \mathcal{P}_{\mathcal{U}} \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2 \sum_{ik} |\mathcal{Z}_{ibk}|^2 + \frac{2}{p} \|\mathbf{T}^{\mathrm{H}}\|^2 \cdot \|\tilde{\mathbf{T}}\|_\infty^2 \sum_{ijk} |\mathcal{Z}_{ijk}|^2 \cdot \|\xi_j^{\mathrm{H}} \star \mathcal{P}_{\mathcal{V}} \star \xi_b\|_{\mathrm{F}}^2 \\
\leq &\frac{2}{p} \|\mathcal{Z}\|_{\infty,2}^2 \max_{ijk} \|T^{\mathrm{H}} \mathcal{P}_{\mathcal{U}} \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2 + \frac{2}{p} \|\mathbf{T}^{\mathrm{H}}\|^2 \cdot \|\tilde{\mathbf{T}}\|_\infty^2 \sum_j \|\xi_j^{\mathrm{H}} \star \mathcal{P}_{\mathcal{V}} \star \xi_b\|_{\mathrm{F}}^2 \sum_{ik} |\mathcal{Z}_{ijk}|^2 \\
\leq &\frac{2}{p} (\max_{ijk} \|T^{\mathrm{H}} \mathcal{P}_{\mathcal{U}} \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2 + \|\mathbf{T}^{\mathrm{H}}\|^2 \cdot \|\tilde{\mathbf{T}}\|_\infty^2 \|\mathcal{P}_{\mathcal{V}} \star \xi_b\|_{\mathrm{F}}^2) \cdot \|\mathcal{Z}\|_{\infty,2}^2 \\
\leq &\frac{2}{p} (\kappa^2(\mathbf{T}) \frac{\nu r}{n_1} + \|\mathbf{T}\|^2 \rho(\mathbf{T}) \frac{\mu r}{n_2}) \|\mathcal{Z}\|_{\infty,2}^2.
\end{aligned} \tag{77}$$

Since $\|\mathbb{E}[\sum_{ijk} \mathtt{vec}(\mathcal{C}_{ijk}) \mathtt{vec}(\mathcal{C}_{ijk})^{\mathrm{H}}]\|$ is the spectral norm of a matrix. When $\{\mathtt{vec}(\mathcal{C}_{ijk})\}_{ijk}$ share the same basis the spectral norm takes the maximum, thus $\|\mathbb{E}[\sum_{ijk} \mathtt{vec}(\mathcal{C}_{ijk}) \mathtt{vec}(\mathcal{C}_{ijk})^{\mathrm{H}}]\| \leq |\mathbb{E}[\sum_{ijk} \mathtt{vec}(\mathcal{C}_{ijk})^{\mathrm{H}} \mathtt{vec}(\mathcal{C}_{ijk})]$.

Using the extension of Theorem A.4:

$$\begin{aligned}
&\|T^{\mathrm{H}} \mathcal{P}_{\mathbb{S}} \tilde{T} (\mathcal{I} - \frac{1}{p} \mathcal{S}_\Omega) \mathcal{Z}\|_{\infty,2} \\
\leq &2c \sqrt{\frac{2(\kappa^2(\mathbf{T}) \frac{\nu r}{n_1} + \|\mathbf{T}\|^2 \rho(\mathbf{T}) \frac{\mu r}{n_2}) \cdot \log(2 \max(n_1, n_2) n_3)}{p}} \|\mathcal{Z}\|_{\infty,2} \\
+ &c \frac{1}{p} \kappa(\mathbf{T}) \sqrt{\frac{\nu r (n_1 + n_2)}{n_1 n_2}} \log(2 \max(n_1, n_2) n_3) \|\mathcal{Z}\|_\infty
\end{aligned} \tag{78}$$

holds with probability at least $1 - (2(n_1 + n_2)n_3)^{1-c}, \forall c > 1$.

The norm of tensor row can be bounded in the same fashion. Thus if $p \geq c_0(\gamma^2(\mathbf{T}) + \rho^2(\mathbf{T})) \frac{\lambda r(n_1+n_2)}{n_1 n_2} \log((n_1 + n_2)N_3)$ with c_0 large enough,

$$\begin{aligned}
\|T^{\mathrm{H}} \mathcal{P}_{\mathbb{S}} \tilde{T} (\mathcal{I} - \frac{1}{p} \mathcal{S}_\Omega) \mathcal{Z}\|_{\infty,2} &\leq 2c \sqrt{\frac{4}{c_0}} \|\mathcal{Z}\|_{\infty,2} + \frac{c}{c_0} \sqrt{\frac{n_1 n_2}{\mu r (n_1 + n_2)}} \|\mathcal{Z}\|_\infty \\
&\leq \frac{1}{2} \|\mathcal{Z}\|_{\infty,2} + \frac{1}{2} \sqrt{\frac{n_1 n_2}{\mu r (n_1 + n_2)}} \|\mathcal{Z}\|_\infty
\end{aligned} \tag{79}$$

holds w.h.p.. □

C Bound Frobenius Norm by Incoherence Condition

Now let us bound $\max_{ijk} \|\mathcal{P}_{\mathbb{S}} T(e_{ijk})\|_{\mathrm{F}}^2$ and $\max_{ijk} \|\mathcal{P}_{\mathbb{S}} \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2$

$$\mathcal{P}_{\mathbb{S}} T(e_{ijk}) = \mathcal{P}_{\mathcal{U}} \star T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}}) + T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}}) \star \mathcal{P}_{\mathcal{V}} - \mathcal{P}_{\mathcal{U}} \star T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}}) \star \mathcal{P}_{\mathcal{V}}, \tag{80}$$

where $\bar{\xi}_i$ and $\bar{\xi}_j^{\mathrm{H}}$ are canonical tensor basis in the original domain.

$$\begin{aligned}&\|\mathcal{P}_{\mathbb{S}}\, T(e_{ijk})\|_{\mathrm{F}}^2\\ =&\langle \mathcal{P}_{\mathcal{U}} \star T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}}),\, T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}})\rangle + \langle\, T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}}) \star \mathcal{P}_{\mathcal{V}},\, T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}})\rangle\\ -&\langle \mathcal{P}_{\mathcal{U}} \star T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}}) \star \mathcal{P}_{\mathcal{V}},\, T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}})\rangle\end{aligned} \tag{81}$$

$$\begin{aligned}\langle \mathcal{P}_{\mathcal{U}} \star T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}}),\, T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}})\rangle =&\langle \mathcal{U}^{\mathrm{H}} \star T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}}), \mathcal{U}^{\mathrm{H}} \star T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}})\rangle\\ =&\|\mathcal{U}^{\mathrm{H}} \star \xi_i \star T(\zeta_k)\|_{\mathrm{F}}^2\end{aligned} \tag{82}$$

In the same way,

$$\begin{aligned}&\langle\, T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}}) \star \mathcal{P}_{\mathcal{V}},\, T(\bar{\xi}_i \star \zeta_k \star \bar{\xi}_j^{\mathrm{H}})\rangle\\ =&\|\mathcal{V}^{\mathrm{H}} \star \xi_j \star T(\zeta_k)\|_{\mathrm{F}}^2 \text{ (here the conjugate does not change the value).}\end{aligned} \tag{83}$$

And we have

$$\|\mathcal{P}_{\mathbb{S}}\, T(e_{ijk})\|_{\mathrm{F}}^2 \le \|\mathcal{U}^{\mathrm{H}} \star \xi_i \star T(\zeta_k)\|_{\mathrm{F}}^2 + \|\mathcal{V}^{\mathrm{H}} \star \xi_j \star T(\zeta_k)\|_{\mathrm{F}}^2. \tag{84}$$

When the energy of $\mathcal{U}(:,:,k)$ is uniformly distributed among its entries, $\max_{ijk} \|\mathcal{U}^{\mathrm{H}} \star \xi_i \star T(\zeta_k)\|_{\mathrm{F}}^2$ takes the minimum as $\frac{r}{n_1}\|\mathbf{T}\|_{1\mapsto 2}^2$; When $\mathcal{U}(:,:,k)$ contains an identity tensor, i.e., the most non-uniformly distributed, $\max_{ijk} \|\mathcal{U}^{\mathrm{H}} \star \xi_i \star T(\zeta_k)\|_{\mathrm{F}}^2$ takes the maximum as $\|\mathbf{T}\|_{1\mapsto 2}^2$. Applying tensor incoherence conditions:

$$\begin{aligned}\|\mathcal{U}^{\mathrm{H}} \star \xi_i \star T(\zeta_k)\|_{\mathrm{F}}^2 &\le \frac{\nu r}{n_1}\|\mathbf{T}\|_{1\mapsto 2}^2\\ \|\mathcal{V}^{\mathrm{H}} \star \xi_j \star T(\zeta_k)\|_{\mathrm{F}}^2 &\le \frac{\nu r}{n_2}\|\mathbf{T}\|_{1\mapsto 2}^2\end{aligned}$$

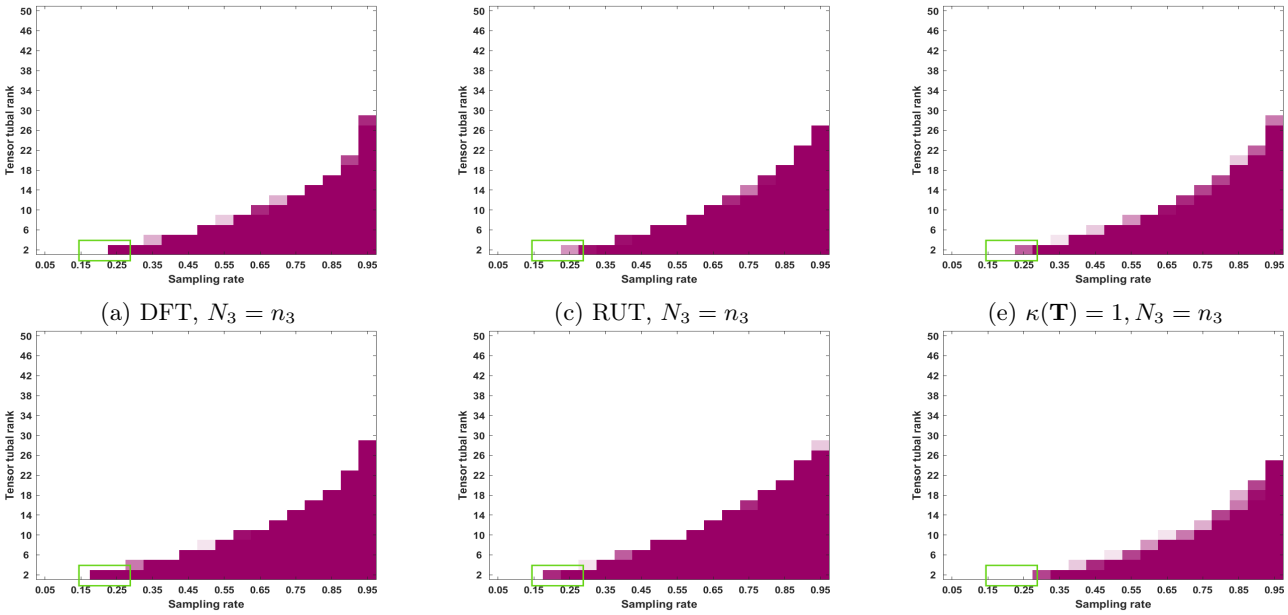
And we have:

$$\|\mathcal{P}_{\mathbb{S}}\, T(e_{ijk})\|_{\mathrm{F}}^2 \le \frac{\nu r}{n_1}\|\mathbf{T}\|_{1\mapsto 2}^2 + \frac{\nu r}{n_2}\|\mathbf{T}\|_{1\mapsto 2}^2 = \frac{\nu r(n_1+n_2)}{n_1 n_2}\|\mathbf{T}\|_{1\mapsto 2}^2 \tag{85}$$

$$\|\mathcal{P}_{\mathbb{S}}\, \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2 \le \frac{\nu r(n_1+n_2)}{n_1 n_2}\|\tilde{\mathbf{T}}\|_{1\mapsto 2}^2 \tag{86}$$

D Additional Experiment Results

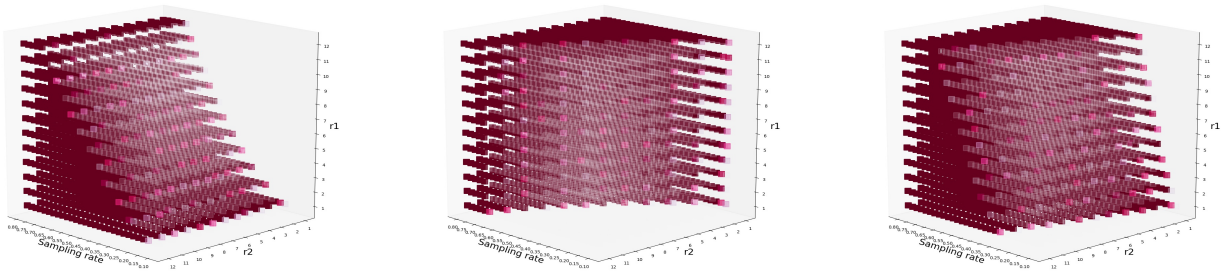
Figure 6 and Figure 7 are the unimposed versions of Figure 3 and Figure 4.



(a) DFT, $N_3 = n_3$ (c) RUT, $N_3 = n_3$ (e) $\kappa(\mathbf{T}) = 1, N_3 = n_3$

(b) DFT, $N_3 = 2n_3$ (d) RUT, $N_3 = 2n_3$ (f) $\kappa(\mathbf{T}) = 4, N_3 = n_3$

Figure 6: Exact completion with varying rank (x-axis) and sampling rate (y-axis). The white cube denotes all failure and the magenta cube denotes all success.



(a) Using RUT1 only (b) Using RUT2 only (c) RUT1 and RUT2 combined

Figure 7: Exact completion with varying rank($T_1(\mathcal{X})$) (z-axis), rank($T_2(\mathcal{X})$) (y-axis) and sampling rate (x-axis). The white cube denotes all failure and the magenta cube denotes all success.

The performance indicators of each test MSI/video in Section 4 for SR = 0.05 are shown in Figure 5/Figure 8.

Figure 9 presents the visual results of MSI completion shown by pseudo-RGB image. It can be observed that the texture and contour are clearer for DCT2, DFT-DCT and Framelet recovered MSIs. Also, the color fidelity is maintained even under a small sampling rate of 0.05.

The visual recovery results of the 30th frame and 60th frame of the video "coastguard" with SR= 0.1 are exhibited in Figure 10. The frames recovered by DCT2, DFT-DCT, Framelet contain fewer noise artifacts, and the margins of the river and the coast are more obvious compared to other methods.

Next we provide statistical significance test results. Since the real-world datasets have a large number of entries, the variances of the experimental results are small and the performances are stable across different sampling instances (also note that the sampling rate lower bound is not specific for each instance).

To illustrate this, we tested the methods in Table 1 on the test MSI "cloth" for 10 random trials. The results

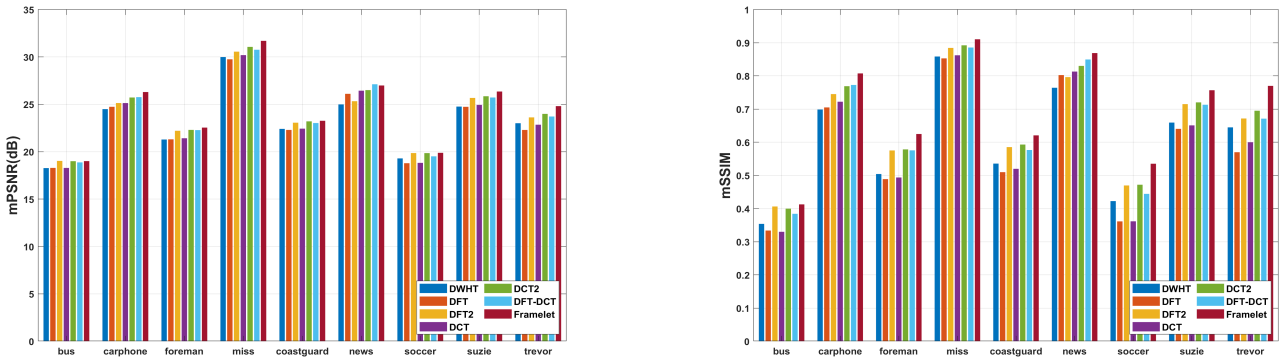
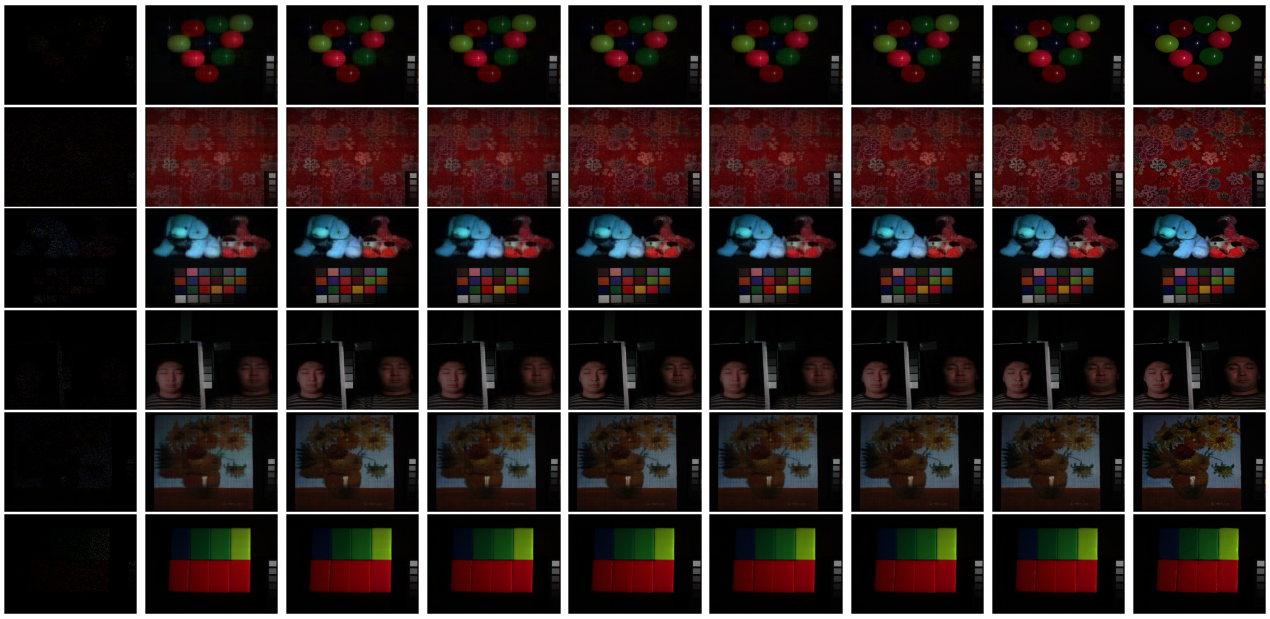


Figure 8: mPSNR (left) and mSSIM (right) results on video data with SR=0.05.



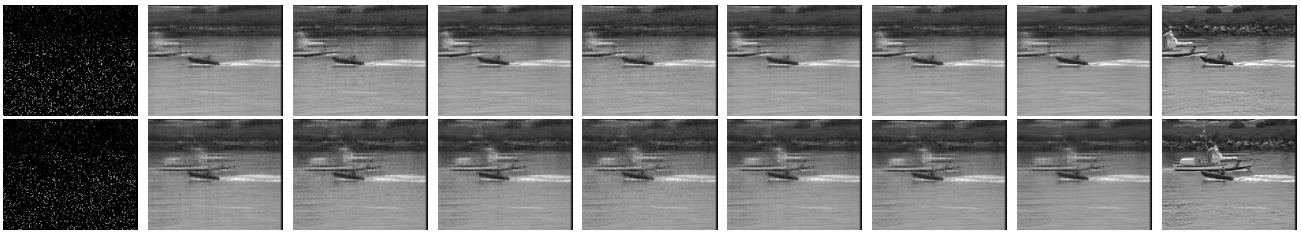
(a) Observed (b) DWHT (c) DFT (d) DFT2 (e) DCT (f) DCT2 (g) DFT-DCT (h) Framelet (i) Original

Figure 9: Pseudo-RGB (R-28, G-15, B-9) image for performance comparison on MSI "super balls", "cloth", "stuffed toys", "photo and face", "oil painting" and "clay" (top to bottom) with SR = 0.05.

with standard deviations are presented in Table 3. We can observe that the performances of these methods are stable and differences between these methods are statistically significant.

We have incorporated cross-paradigm baseline methods, including Bayesian CP Factorization (BCPF) (Zhao et al., 2015), Tucker (Liu et al., 2013), Tensor Train (TT) (Oseledets, 2011), Latent Space Tensor Ring Low-Rank Factors (TRLRF) (Yuan et al., 2019), and Tensor Completion by Parallel Matrix Factorization via TT (TMac-TT) (Bengua et al., 2017). These methods collectively span a set of distinct tensor modeling assumptions: BCPF, Tucker, TT, and TRLRF capture different intrinsic tensor structures, while TMac-TT and KBR represent modern non-convex approaches to tensor completion. Experiments are conducted on visual datasets: MSI and video data, consistent with the experiment section. The results are presented in Table 4 below.

Consistent with the advantages of slim transforms reported in prior work (Jiang et al., 2020; Wang et al., 2021), the tensor completion methods with slim transforms in Table 1 of the manuscript generally outperform the baseline methods in Table 4. The only exception is BCPF, which achieves higher PSNR than DFT2, DCT2, and DFT-DCT at the sampling rate $p = 0.05$; however, it still underperforms Framelet, and for SSIM as well as all other sampling rates, FBCP remains inferior to all slim-transform-based methods.



(a) Observed (b) DWHT (c) DFT (d) DFT2 (e) DCT (f) DCT2 (g) DFT-DCT (h) Framelet (i) Original

Figure 10: Performance Comparison on Video "Coastguard" with SR= 0.1: The first row shows the 30th frame and the second row shows the 60th frame.

Table 3: Mean and standard deviation of the performance (PSNR and SSIM) of the methods in Table 1 on the test MSI "cloth" over 10 random trials.

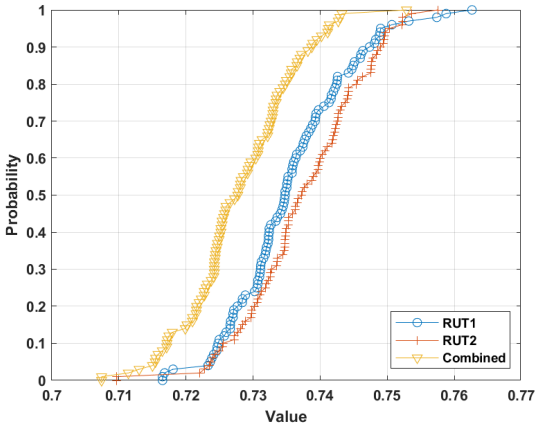
		DWHT	**DFT**	**DFT2**	**DCT**	**DCT2**	**DFT-DCT**	**Framelet**
PSNR	mean	26.0153	26.7200	26.9175	27.7495	27.6549	28.3202	29.3797
	STD	0.0151	0.0191	0.0183	0.0170	0.0182	0.0182	0.0164
SSIM	mean	0.6248	0.6765	0.6944	0.7223	0.7375	0.7662	0.8194
	STD	9.7169e-04	0.0011	0.0011	8.4023e-04	9.3148e-04	8.3305e-04	7.3248e-04

Table 4: Performances of Additional Baselines for MSI and Video Datasets.

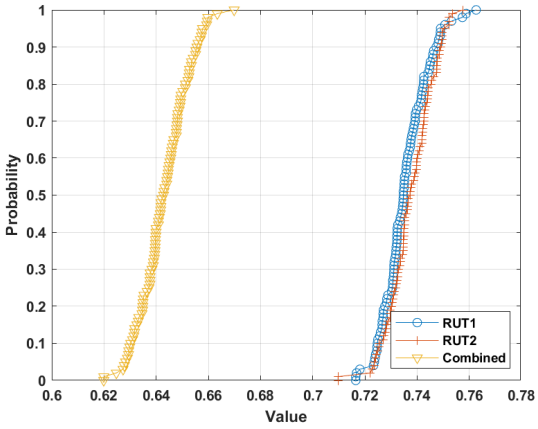
Data	SR	Tucker	TT	TRLRF	BCPF	TMacTT
MSI	p=0.05	27.916/0.7937	26.993/0.7898	30.167/0.7216	34.036/0.8312	32.009/0.8142
	p=0.1	32.522/0.8691	32.184/0.8729	32.414/0.8061	35.643/0.8640	35.231/0.8771
	p=0.15	35.589/0.9111	35.559/0.9181	33.342/0.8303	36.393/0.8753	36.225/0.8940
Video	p=0.05	18.812/0.5120	19.131/0.5248	22.243/0.5672	22.374/0.5826	22.010/0.5191
	p=0.1	21.266/0.6006	21.655/0.6135	32.334/0.6319	23.580/0.6265	24.532/0.6501
	p=0.15	23.078/0.6727	23.468/0.6814	24.010/0.6643	24.319/0.6550	25.682/0.7022

E Analyzing the Overall Sampling Rate

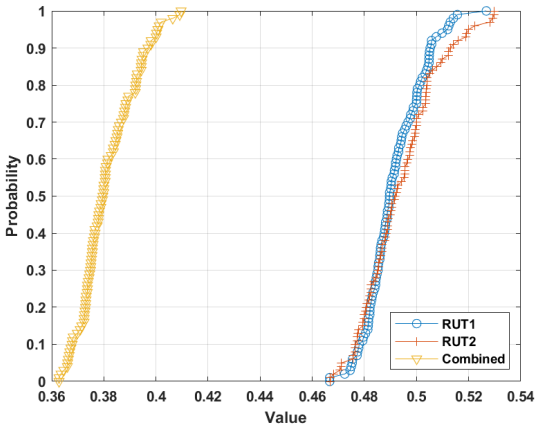
To arrive at a relatively concise form of Theorem 3.4, several Scaling techniques have been used (e.g. the introduction of incoherence parameters), which partially hinders our insights to see which terms are affecting the sampling rate p. In this section, removing redundant (while formally useful) scalings, we directly study the functional terms through the construction of the dual certificate A.4.



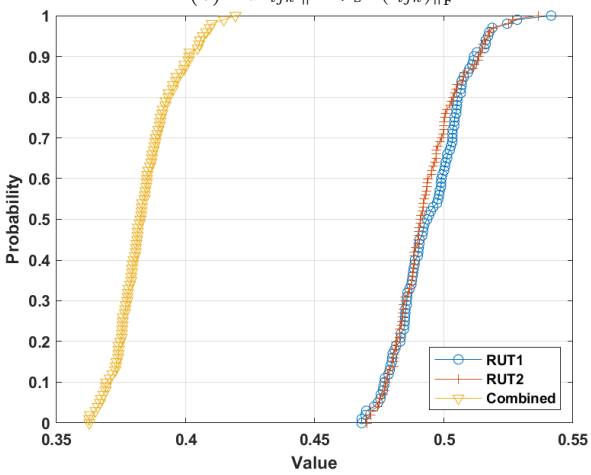
(a) $\gamma(\mathbf{T}) \cdot \max_{ijk}\{\|\mathcal{P}_{\mathbb{S}}\, T(e_{ijk})\|_{\mathrm{F}}^2 \vee \|\mathcal{P}_{\mathbb{S}}\, \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2\}$



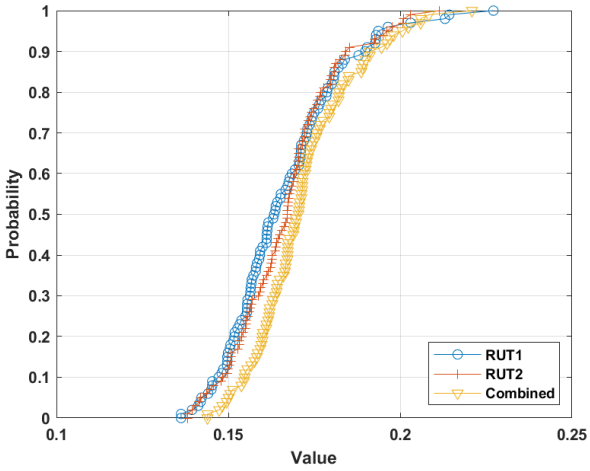
(b) $\max_{ijk} \| T^{\dagger}\mathcal{P}_{\mathbb{S}}\, T(e_{ijk})\|_{\mathrm{F}}^2$



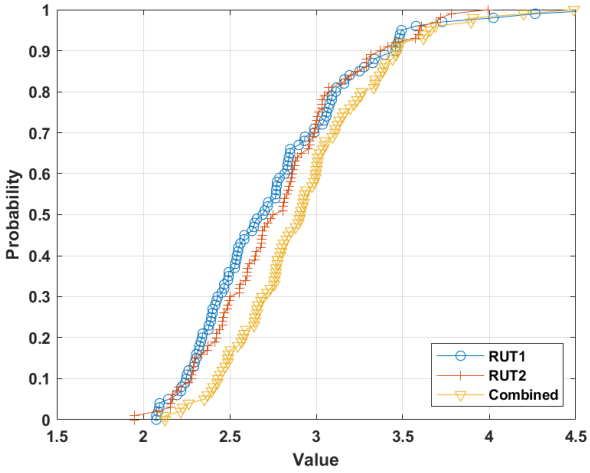
(c) $\max_{ijk} \| T^{\mathrm{H}}\mathcal{P}_{\mathcal{U}}\, \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2$



(d) $\max_{ijk} \| T^{\mathrm{H}}\mathcal{P}_{\mathcal{V}}\, \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2$



(e) $\|\tilde{\mathbf{T}}\|_{\infty} \cdot \| T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\infty}$



(f) $\|\tilde{\mathbf{T}}\|_{\infty}^2 \cdot \| T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\infty,2}^2$

Figure 11: Emperical cumulative distribution of the factors in 94.

From the proof of Proposition A.3, one can observe that the satisfaction of the conditions of dual certificate 6, 7, 8 and 9 put requirements on the sampling rate p. Among them, 7 is satisfied by the structure of the dual certificate. 6 and 8 are essentially the same, only that 8 is harsher since it is upon the step sampling rate p_i ($p = l \cdot p_i$). Thus, we only need to study the requirements posed by 8 and 9.

- To satisfy 8: by 42, one can see that p_i is propotional to the energy terms:

$$p_i \sim \gamma(\mathbf{T}) \cdot \max_{ijk} \|\mathcal{P}_{\mathbb{S}} T(e_{ijk})\|_{\mathrm{F}}^2 \vee \|\mathcal{P}_{\mathbb{S}} \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2,$$
$$p_i \sim \max_{ijk} \|\mathcal{P}_{\mathbb{S}} \tilde{T}(e_{ijk})\|_{\mathrm{F}} \cdot \max_{ijk} \|\mathcal{P}_{\mathbb{S}} T(e_{ijk})\|_{\mathrm{F}}.$$

  Since $\max_{ijk} \|\mathcal{P}_{\mathbb{S}} \tilde{T}(e_{ijk})\|_{\mathrm{F}} \cdot \max_{ijk} \|\mathcal{P}_{\mathbb{S}} T(e_{ijk})\|_{\mathrm{F}} \leq \frac{1}{2}(\max_{ijk} \|\mathcal{P}_{\mathbb{S}} \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2 + \max_{ijk} \|\mathcal{P}_{\mathbb{S}} T(e_{ijk})\|_{\mathrm{F}}^2)$, thus the satisfaction of 8 only requires

$$p_i \sim \gamma(\mathbf{T}) \cdot \max_{ijk} \|\mathcal{P}_{\mathbb{S}} T(e_{ijk})\|_{\mathrm{F}}^2 \vee \|\mathcal{P}_{\mathbb{S}} \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2. \tag{87}$$

- To satisfy 9: there are three lemmas used when proving 9. We need to study the requirements posed by the assumptions of Lemma A.6 and A.7 (Lemma A.5 already holds).
  - To satisfy the assumption of Lemma A.6: by 70, we can observe that:

$$p_i \sim \max_{ijk} \| T^{\dagger} \mathcal{P}_{\mathbb{S}} T(e_{ijk})\|_{\mathrm{F}}^2, \tag{88}$$
$$p_i \sim \max_{ijk} \|\mathcal{P}_{\mathbb{S}} \tilde{T}(e_{ijk})\|_{\mathrm{F}} \cdot \max_{ijk} \|\mathcal{P}_{\mathbb{S}} T(e_{ijk})\|_{\mathrm{F}}. \tag{89}$$

  - To satisfy the assumption of Lemma A.7: there are two terms affecting p_i in this lemma: $\max_{ijk} \| T^{\mathrm{H}} \mathcal{P}_{\mathcal{U}} \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2$ and $\|\mathbf{T}\|^2 \|\tilde{\mathbf{T}}\|_{\infty}^2 \|\mathcal{P}_{\mathcal{V}} \star \xi_b\|_{\mathrm{F}}^2$, both coming from $\max_{ijk} \| T^{\mathrm{H}} \mathcal{P}_{\mathcal{U}} \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2$. The reason why for slim transforms the first term is smaller is the same as illustrated in Section 3.5. For the second term, which has been scaled to introduce incoherence parameters, since we require no structures on $\mathbf{T}$, it can not be simplified further to offer more insights. However, we can still directly analyze it: for a combinational transform $\mathbf{T}_{\mathrm{c}} = \frac{1}{\sqrt{n}}[\mathbf{T}_1^{\mathrm{T}}, \mathbf{T}_2^{\mathrm{T}}, \ldots, \mathbf{T}_M^{\mathrm{T}}]^{\mathrm{T}}$, $\|\mathbf{T}_{\mathrm{c}}\|^2 \|\tilde{\mathbf{T}}_{\mathrm{c}}\|_{\infty}^2 = \max\left\{\|\mathbf{T}_m\|^2 \|\tilde{\mathbf{T}}_m\|_{\infty}^2\right\}$; also $\|\mathcal{P}_{\mathbb{S}_{\mathrm{c}}} \star \xi_b\|_{\mathrm{F}}^2 = \mathrm{mean}\left\{\|\mathcal{P}_{\mathbb{S}_m} \star \xi_b\|_{\mathrm{F}}^2\right\}$. Thus, when $\mathbf{T}_m$ contains not too less entries, $\|\mathbf{T}\|^2 \|\tilde{\mathbf{T}}\|_{\infty}^2 \|\mathcal{P}_{\mathbb{S}} \star \xi_b\|_{\mathrm{F}}^2$ is approximately the same for slim and square transforms. Therefore, the differentiating requirements would be:

$$p_i \sim \max_{ijk} \| T^{\mathrm{H}} \mathcal{P}_{\mathcal{U}} \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2, \tag{90}$$
$$p_i \sim \max_{ijk} \| T^{\mathrm{H}} \mathcal{P}_{\mathcal{V}} \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2. \tag{91}$$

  - To satisfy the requirements posed by $\| T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\infty}$ and $\| T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\infty,2}^2$: to finally satisfy 9, the requirements are:

$$p_i \sim \|\tilde{\mathbf{T}}\|_{\infty} \cdot \| T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\infty}, \tag{92}$$
$$p_i \sim \|\tilde{\mathbf{T}}\|_{\infty}^2 \cdot \| T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\infty,2}^2. \tag{93}$$

Thus, the sampling rate is proportional to the following terms:

$$p \sim \begin{cases} \gamma(\mathbf{T}) \cdot \max_{ijk}(\|\mathcal{P}_{\mathbb{S}}\, T(e_{ijk})\|_{\mathrm{F}}^2 \vee \|\mathcal{P}_{\mathbb{S}}\, \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2) \\ \max_{ijk} \|\, T^{\dagger}\mathcal{P}_{\mathbb{S}}\, T(e_{ijk})\|_{\mathrm{F}}^2 \\ \max_{ijk} \|\, T^{\mathrm{H}}\mathcal{P}_{\mathcal{U}}\, \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2 \\ \max_{ijk} \|\, T^{\mathrm{H}}\mathcal{P}_{\mathcal{V}}\, \tilde{T}(e_{ijk})\|_{\mathrm{F}}^2 \\ \|\tilde{\mathbf{T}}\|_{\infty} \cdot \|\, T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\infty}, \\ \|\tilde{\mathbf{T}}\|_{\infty}^2 \cdot \|\, T^{\mathrm{H}}(\mathcal{U} \star \mathcal{V}^{\mathrm{H}})\|_{\infty,2}^2. \end{cases} \tag{94}$$

We use the same settings in Section 3.5 and conducted the Monte-Carlo experiment for 100 times to plot the empirical cumulative distribution of the terms in 94 for two random unitary transforms RUT1 and RUT2, and the combined transform.

The distributions of these terms have been shown in Figure 11. One can observe that, except that in Figure 11e and 11f the factors of the slim transform are marginally larger than those of the square transforms, the rest terms of the slim transform are all distributed in a region of smaller values, which results in a smaller sampling rate that we need to achieve exact completion.

To prove Theorem 3.7, one can note that the sampling rate lower bound $p(\mathbf{T})$ is composed of a linear factor and a log factor. In the region near the minimum sampling rate, the linear factor represented by $E_2(\mathbf{T})$ decreases with N_3 (almost inversely proportional to N_3 within a certain range, e.g., $N_3 \in [n_3, 2n_3]$). For the log factor, i.e., $\log^2\big(\gamma(\mathbf{T})(n_1+n_2)N_3\big)$, one can note that due to the existence of n_1, n_2, N_3, the term in the log function is large, thus having entered the growth-flat region of the log function ($\frac{\log((n_1+n_2)N_3)}{\log((n_1+n_2)n_3)} \approx 1$) and diminishes in the ratio in Theorem 3.7.

F Algorithm to solve Program 1

Program 1 can be solved via standard ADMM. By introducing the auxiliary variables $\mathcal{Y}, \mathcal{Z} \in \mathbb{C}^{n_1 \times n_2 \times N_3}$, the augmented Lagrangian can be written as:

$$\mathcal{L}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) = \min_{\mathcal{X},\mathcal{Y},\mathcal{Z}} \|\mathcal{Y}\|_* + \langle T(\mathcal{X}) - \mathcal{Y}, \mathcal{Z}\rangle + \frac{\alpha}{2}\|\, T(\mathcal{X}) - \mathcal{Y}\|_{\mathrm{F}}^2 + \mathcal{I}_{\Omega}(\mathcal{X} - \mathcal{M}) \tag{95}$$

where $\mathcal{I}_{\Omega}(\mathcal{A})$ equals to 0 if $\mathcal{P}_{\Omega}(\mathcal{A}) = \mathbf{0}$ and equals to $+\infty$ elsewhere.

- Solving $\mathcal{Y}$ sub-problem:

$$\mathcal{Y} = \arg\min_{\mathcal{Y}} \|\mathcal{Y}\|_* + \frac{\alpha}{2}\|\, T(\mathcal{X}) - \mathcal{Y} + \frac{\mathcal{Z}_1}{\alpha}\|_{\mathrm{F}}^2$$

  Similar to (Lu et al., 2019) while being more intuitive in form since the tensor nuclear norm is directly defined on the transform domain, define $\mathcal{D}_{\tau}(\cdot)$ as the singular value thresholding operator and it is the proximity operator of 96:

$$\mathcal{D}_{\tau}(\mathcal{A}) = \arg\min_{\mathcal{X}} \tau\|\mathcal{X}\|_* + \frac{1}{2}\|\mathcal{X} - \mathcal{A}\|_{\mathrm{F}}^2 = \mathcal{U} \star (\mathcal{S} - \tau)_+ \star \mathcal{V}^{\mathrm{H}}, \tag{96}$$

  where $\mathcal{A} = \mathcal{U} \star \mathcal{S} \star \mathcal{V}^{\mathrm{H}}$.

  Using $\mathcal{D}_{\tau}(\cdot)$, $\mathcal{Y}$ can be computed as

$$\mathcal{Y}^{t+1} = \mathcal{D}_{1/\alpha^t}(\, T(\mathcal{X}^t) + \frac{\mathcal{Z}^t}{\alpha^t}) \tag{97}$$

- Solving $\mathcal{X}$ sub-problem:

$$\mathcal{X}^{t+1} = \arg\min_{\mathcal{X}} \frac{\alpha}{2}\|\, T(\mathcal{X}) - \mathcal{Y} + \frac{\mathcal{Z}_1}{\alpha}\|_{\mathrm{F}}^2 + \mathcal{I}_{\Omega}(\mathcal{X} - \mathcal{M}) \tag{98}$$

$$= \mathcal{S}_{\Omega^{\mathrm{C}}}(\, T^{\dagger}(\mathcal{Y}^t - \frac{\mathcal{Z}^t}{\alpha^t})) + \mathcal{S}_{\Omega}(\mathcal{M}) \tag{99}$$

- Computing Lagrangian multiplier

$$\mathcal{Z}^{t+1} = \mathcal{Z}^t + \alpha^t(T(\mathcal{X}^{t+1}) - \mathcal{Y}^{t+1}) \tag{100}$$

The whole process is given in Algorithm 2.

Algorithm 2 Solving 1 by ADMM

Input: $\mathcal{X}_0 \in \mathbb{C}^{n_1 \times n_2 \times n_3}, \mathbf{T} \in \mathbb{C}^{N_3 \times n_3}, \Omega, \rho, \alpha_0, \alpha_{\max}$
Output: $\mathcal{X} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$
Initialize: $\mathcal{X} = \mathcal{X}_0, \mathcal{Y} = T(\mathcal{X}), \mathcal{Z} = \mathbf{0}, \alpha = \alpha_0$
while no convergence **do**
1. Compute $\mathcal{Y}^{t+1}$ by eq. 97;
2. Compute $\mathcal{X}^{t+1}$ by eq. 99;
3. Compute $\mathcal{Z}^{t+1}$ by eq. 100;
4. $\alpha^{t+1} = \min\{\rho\alpha^t, \alpha_{\max}\}$;

end while

G Generate $\mathcal{M}$ by Alternating Projection

Algorithm 3 Generate $\mathcal{M}$ with a specified rank($T(\mathcal{M})$)

Input: $\mathcal{M}_0 \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, $T : \mathbb{C}^{n_3} \mapsto \mathbb{C}^{N_3}, r$
Output: $\mathcal{M} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$
Initialize: $\mathcal{M} = \mathcal{M}_0$
while no convergence **do**
1. Project $T(\mathcal{M})$ onto a tubal rank-r manifold using truncated t-SVD and obtains $\mathcal{N}$;
2. Project $\mathcal{N}$ back to the column space spanned by T: $T(\mathcal{M}) = T(T^\dagger(\mathcal{N}))$;
3. $T(\mathcal{M}) = T(\mathcal{M})/\|T(\mathcal{M})\|_{\mathrm{F}}$;

end while

In this section, we introduce the alternating projection algorithm to generate $\mathcal{M}$ with a specified rank($T(\mathcal{M})$). First, provided with a randomly initiated tensor $\mathcal{M}_0$ and a preassigned transform $T(\cdot)$, we project $T(\mathcal{M})$ onto a low tubal rank manifold using truncated t-SVD. In matrix case, the truncated SVD gives the projection onto nearest low rank matrix manifold in the sense of the metric induced by the Frobenius norm. The truncated t-SVD also enjoys such property following the same reason. After the projection onto low tubal rank manifold, we project $T(\mathcal{M})$ onto the column space spanned by $\mathbf{T}$ to ensure the preimage $\mathcal{M}$ exists via least square, which also minimizes the difference of the energy before and after the projection (measured by the Frobenius norm). Also we keep the energy of $T(\mathcal{M})$ to avoid iterating to zero. By repeating these three steps we can generated random tensors $\mathcal{M}$ with a specified rank($T(\mathcal{M})$) and the procedure is given in Algorithm 3.

Algorithm 4 Generate $\mathcal{M}$ with specified rank($T_1(\mathcal{M})$) and rank($T_2(\mathcal{M})$)

Input: $\mathcal{M}_0 \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, $T_1 : \mathbb{C}^{n_3} \mapsto \mathbb{C}^{M_3}, r_1$, $T_2 : \mathbb{C}^{n_3} \mapsto \mathbb{C}^{N_3}, r_2$
Output: $\mathcal{M} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$
Initialize: $\mathcal{M} = \mathcal{M}_0$
while no convergence **do**
1. Project $T_1(\mathcal{M})$ onto a tubal rank-r manifold using truncated t-SVD and obtains $\mathcal{N}_1$;
2. Project $\mathcal{N}_1$ back to the preimage space of the column space spanned by T_1: $\mathcal{M} = T_1^\dagger(\mathcal{N}_1)$;
3. $\mathcal{M} = \mathcal{M}/\|\mathcal{M}\|_{\mathrm{F}}$;
4. Project $T_2(\mathcal{M})$ onto a tubal rank-r manifold using truncated t-SVD and obtains $\mathcal{N}_2$;
5. Project $\mathcal{N}_2$ back to the preimage space of the column space spanned by T_2: $\mathcal{M} = T_2^\dagger(\mathcal{N}_2)$;
6. $\mathcal{M} = \mathcal{M}/\|\mathcal{M}\|_{\mathrm{F}}$;

end while

In the experiments, for performance comparison, generating $\mathcal{M}$ with rank($T_1(\mathcal{M})$) $= r_1$ and rank($T_2(\mathcal{M})$) $= r_2$ is needed. Proceeding from Algorithm 3, an algorithm with two sets of projections is designed in Algorithm 4. Note that in Algorithm 4, since it is $\mathcal{M}$ being operated thus we need to project $\mathcal{N}_1$ and $\mathcal{N}_2$ onto the preimage space.

H Complexity

Though enjoying many satisfying properties, one frequently mentioned issue for the approach using nuclear norm minimization is the scaling problem since it usually requires time-consuming SVD operation. In this paper, the problem can be alleviated with the technique of parallelization.

The computational complexity of the nuclear norm minimization algorithm in the paper is $\mathcal{O}(\max(n_1, n_2) \cdot \min^2(n_1, n_2)N_3 + n_1 n_2 n_3 N_3)$ per iteration for all transforms. The first part corresponds to the truncated t-SVD operation, which involves computing SVD with complexity $\mathcal{O}(\max(n_1, n_2) \cdot \min^2(n_1, n_2))$ for all N_3 slices. It can be parallelized such that SVD of all N_3 slices can be calculated simultaneously, which mitigates the external running time burden brought by slim transforms. The second part refers to the domain transform operation.