Recovering low-rank tensor from limited coefficients in any ortho-normal basis using tensor-singular value decomposition

Shuli Ma\(^1,2\)
Jianhang Ai\(^3\)
Huiqian Du\(^1\)
Liping Fang\(^4\)
Wenbo Mei\(^1\)

\(^1\)School of Information and Electronics, Beijing Institute of Technology, Beijing, China
\(^2\)School of Space Information, Space Engineering University, Beijing, China
\(^3\)Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic
\(^4\)School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China

Abstract
Tensor singular value decomposition (t-SVD) provides a novel way to decompose a tensor. It has been employed mostly in recovering missing tensor entries from the observed tensor entries. The problem of applying t-SVD to recover tensors from limited coefficients in any given ortho-normal basis is addressed. We prove that an \(n \times n \times n\) tensor with tubal-rank \(r\) can be efficiently reconstructed by minimising its tubal nuclear norm from its \(O(mn)\) \(\log^2(n)\) randomly sampled coefficients w.r.t any given ortho-normal basis. In our proof, we extend the matrix coherent conditions to tensor coherent conditions. We first prove the theorem belonging to the case of Fourier-type basis under certain coherent conditions. Then, we prove that our results hold for any ortho-normal basis meeting the conditions. Our work covers the existing t-SVD-based tensor completion problem as a special case. We conduct numerical experiments on random tensors and dynamic magnetic resonance images (d-MRI) to demonstrate the performance of the proposed methods.

1 INTRODUCTION

Using tensors and their decompositions to process data has received great attention in many fields, such as computer vision, signal/image processing, data mining etc. An important observation in applications is that tensors often possess the low-rank property/structure even though they reside in extremely high-dimensional data spaces. Consequently, various tensor decomposition frameworks have been developed to better represent the inner structure of a tensor such as CANDECOMP/PARAFAC (CP), Tucker decomposition \([1, 2]\) and recently proposed a tensor train (TT) \([3]\), a tensor ring (TR) \([4, 5]\) as well as a tensor-singular value decomposition (t-SVD) \([6–9]\). An overview of the tensor decompositions for multi-way visual data completion can be found in ref. \([2]\). The commonly used method for calculating CP decomposition is the alternating least squares (ALS) method. Paper \([10–12]\) used the computationally expensive iterative ALS algorithm to solve the tensor PARAFAC decomposition model. But the ALS method cannot guarantee convergence to the global minimum. The final solution can heavily depend on the initial guess as well \([1]\). Therefore, in CP decomposition, CP rank is hard to be estimated \([1]\). In Tucker decomposition, Tucker rank is composed of every rank of mode-n matrices and these matrices are extremely unbalanced (the number of columns is much larger than that of the rows, or vice versa). In TT decomposition, TT rank consists of ranks of matrices formed by a well-balanced matricization scheme and is more efficient for high-order tensors. TR is a general-purpose network derived from TT, and in special cases, the two are equal. Among all these decompositions, t-SVD is more suitable for order-3 tensor decomposition, which is similar to the SVD for matrices. It can extend matrix analysis to the multi-linear setting without the loss of inherent information present in tensors. T-SVD has shown good performance in tensor completion \([13, 14]\), data de-noising \([7]\), image de-blurring \([9]\), tensor robust principal component analysis \([7, 15]\) etc.

Generally, researchers enforce the low rankness of the tensor in low-rank tensor completion (LRTC), which aims to recover missing entries of a tensor from its partially observed entries. T-SVD has been recently exploited for LRTC \([6, 7, 13]\).
Chen et al. [16] proposed a LRTC method via t-SVD for image and video recovery. Zhang and Aeron [6] proved (and then Lu et al [14] corrected) that it is possible to exactly complete a tensor with high probability from the $O(\tau \min\{n_1, n_2, n_3\} \log^2\min\{n_1, n_2, n_3\})$ uniformly sampled elements by enforcing the low tubal rank of it, given the tensor size is $n_1 \times n_2 \times n_3$ and tubal-rank is $r$.

In some applications such as synthetic aperture radar (SAR) imaging [17, 18], dynamic magnetic resonance imaging (d-MRI) [19, 20], a limited number of measurements, which are expansion coefficients with respect to some basis, are obtained to reconstruct the original images. For example, sparse sampling in MRI can accelerate imaging speed significantly [16, 21–25]. In recent years, we have tried to use different tensor decomposition frameworks in sparse sampling MRI. In our previous work [21], the TT decomposition framework is used to improve the quality of dynamic MR image reconstruction under sparse sampling. In ref. [22], we combined the partially separable scheme with tensor low tubal rank constraint to further reduce sampling data. In the conference paper [25], we gave the result of using t-SVD model and the sparsity for MRI reconstruction. Different from ref. [25], we focus on theoretical analysis and proof, and give the theorem and coherence conditions for the tensor reconstruction from limited coefficients in any orthonormal basis. In addition, we demonstrate the experimental results of the proposed method and its variants substantially.

We address the problem of applying t-SVD to reconstruct tensors from limited coefficients in any given orthonormal basis. The tensor recovery problem we focus on can be seen as a generalisation of the tensor completion problem in ref. [6]. With our theoretical support, the prior low tubal rank constraint on t-SVD framework can not only be applied to video completion problems like Zhang and Aeron did in [6, 7], but also can be applied to other applications. We propose a tensor reconstruction model based on t-SVD and derive theoretical performance bounds for it. The main theoretical result is Theorem 2 in Section 3. Specifically, we extend the matrix coherent conditions in [26] to tensor coherent conditions and prove that our results hold for any orthonormal basis meeting the conditions. In the proof, we first validate our results in the case of Fourier-type basis. Then, we generalise the proof to any orthonormal basis using mathematical induction with two modifications. Similar to ref. [6], we also have to handle the block diagonal constraint in the proof procedure, which leads to subtle differences with matrix reconstruction problems in ref. [26]. We adopted the Golging scheme1 in our proof which is widely used after [26] first introduced. The difference is in our generalisation, the critical step is how to verify that the condition $(2, b)$ in Proposition 1 still holds under any orthonormalised basis. Lastly, we conduct numerical experiments (including dynamic MR image reconstruction) to demonstrate the performance of the proposed methods.

The remainder is organized as follows. In Section 2, t-SVD is reviewed briefly. In Section 3, the main theoretical result is given. In Section 4, the proofs of the main theoretical result are provided. In Section 5, the experimental results are presented to demonstrate the effectiveness of the proposed method. In Section 6, the conclusion is made. In Section 7, the appendices is present.

2 | BRIEF OVERVIEW OF t-SVD

2.1 | Notations and preliminaries

A summary of symbols and notations is shown in Table 1.

In the scenario of tensor decomposition, a slice is defined as a matrix obtained by fixing every index but two indexes of a tensor. A fibre is defined by fixing every index but one. For an order-3 tensor $\mathcal{A}$, we use the MATLAB notation $\mathcal{A}(\mathbf{k}, :, :)$, $\mathcal{A}(; k, :)$, $\mathcal{A}(; :, k)$ and $\mathcal{A}(; :, :, k)$ to denote the $k_{\text{th}}$ horizontal, lateral and frontal slices, and $\mathcal{A}(; i, j)$, $\mathcal{A}(i, :, j)$ and $\mathcal{A}(i, j, :)$ to denote the $(i, j)_{\text{th}}$ mode-1, mode-2, mode-3 fibre. We use $\mathcal{A} = \text{fft}([\mathcal{A}(:, :) \mathcal{A}(::), \mathcal{A}(::)]$ as the tensor obtained by applying the 1D-FFT along the third dimension of $\mathcal{A}$. Especially, we use $\mathcal{A}^{(k)}$ to represent $\mathcal{A}(; : , k)$, and use $\widehat{\mathcal{A}}^{(k)}$ to represent $\widehat{\mathcal{A}}(; : , k)$. Let $\mathcal{A}$ denote the block-diagonal matrix of the tensor $\mathcal{A}$ in the Fourier domain, as shown in Equation (1).

The following preliminaries will be used frequently.

1. Norms of Matrix $M$. 

| TABLE 1 | Basic notations |
|---|---|
| Symbols | Notations |
| $a, \mathcal{A}$ | Scalar, tensor (order $\geq 3$) |
| $\mathcal{A}(i_1, i_2, \ldots , i_N)$ | The element of $N$-order tensor $\mathcal{A}$ |
| $\mathcal{A}(; : , k)$ or $\mathcal{A}^{(k)}$ | The $k_{\text{th}}$ frontal slice of a 3-order $\mathcal{A}$ |
| $\mathcal{A}(; : , k)$ | The vector reshaped from $\mathcal{A}$ |
| Superscript $H$ | Hermitian transpose |
| Superscript $T$ | Tensor transpose |
| Top mark $\sim$ | 1D-FFT along the third dimension |
| Top mark $\bigcirc$ | Block-diagonal matrix in the Fourier domain |
| $\circ$ | Hadamard product |
| $*$ | t-product |

---

1. Here, we use $W_{x}(; : )$ to represent the vector size of $n_{1}n_{2} \times n_{3}$ which is reshaped from $W_{x}$. Then, the element $W_{x}(a, j, k)$ in $W_{x}$ is the corresponding $a_{b}$ element of vector $W_{x}(; : )$, where $a = 1, 2, \ldots , n_{1}n_{2}n_{3}$. 

Operator norm $\|M\|_{op}$ or spectral norm $\|M\|$: the largest singular value of $M$.

The Frobenius norm $\|M\|_F$: trace$(M,M)^{\frac{1}{2}} = (\sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij}^2)^{\frac{1}{2}}$.

Nuclear norm $\|M\|_*$: the sum of the singular values of matrix $M$.

2. Norms of Tensor $\mathbf{A}$.

Operator norm $\|\mathbf{A}\|_{op}$ or spectral norm $\|\mathbf{A}\|$: the largest singular value of $\mathbf{A}$, and $\|\mathbf{A}\|_{op} = \|\mathbf{A}\|_{ap}$.

Frobenius norm $\|\mathbf{A}\|_F$: $(\mathbf{A},\mathbf{A})^{\frac{1}{2}} = \sum_{i,j,k} A_{ijk}^2$.

Tensor nuclear norm $\|\mathbf{A}\|_{TNN}$: the sum of singular values of all frontal slices of $\mathbf{A}$, and $\|\mathbf{A}\|_{TNN} = \|\mathbf{A}\|_*$.

3. Two inequalities for any same size matrices $\mathbf{A}$ and $\mathbf{B}$:

$$\langle A, B \rangle \leq \|A\|_F \|B\|_F$$

$$\langle A, B \rangle \leq \|A\| \|B\|_*$$

### 2.2 Tensor singular value decomposition

The introduction to t-SVD below is for the reader to understand our main result, more information about t-SVD can be found in refs. [6, 9, 27–29].

**Definition 1** t-product [29]. The t-product of $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathbf{B} \in \mathbb{R}^{n_2 \times n_1 \times n_3}$ is a tensor of size $n_1 \times n_2 \times n_3$

$$\mathbf{A} \ast \mathbf{B} = \mathbf{C} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$$

where $\ast$ denotes t-product, $\mathbf{C}(i,j,:)$ is given by $\sum_{k=1}^{n_2} \mathbf{A}(i,k,:) \circ \mathbf{B}(k,j,:)$ among which $\circ$ denotes the circular convolution between the two vectors, and $i = 1, 2, ..., n_1$, $j = 1, 2, ..., n_4$.

**Definition 2** Tensor transpose [28, 29]. We denote the transpose of an $n_1 \times n_2 \times n_3$ tensor $\mathbf{A}$ as $\mathbf{A}^T$ of size $n_2 \times n_1 \times n_3$. $\mathbf{A}^T$ can be obtained from $\mathbf{A}$ by transposing each of the frontal slices first, then reversing the order of transposed frontal slices 2 through $n_3$.

$$\begin{align*}
(\mathbf{A}^T)^{(1)} &= (\mathbf{A}^{(1)})^T \\
(\mathbf{A}^T)^{(i)} &= (\mathbf{A}^{(n_2+i-1)})^T, \quad i = 2, \ldots, n_3
\end{align*}$$

**Definition 3** Identity tensor. Tensor $\mathbf{J}$ size of $n \times n$ is called identity tensor if the frontal slice $\mathbf{J}^{(1)}$ is the $n \times n$ identity matrix and all other frontal slices $\mathbf{J}^{(i)}$, $i = 2, \ldots, n_3$ are zero.

**Definition 4** Orthogonal tensor. Tensor $\mathbf{A}$ size of $n \times n \times n_3$ is called orthogonal tensor if $\mathbf{A} \ast \mathbf{A}^T = \mathbf{J}$, where $\mathbf{J}$ is identity tensor.

**Definition 5** f-diagonal tensor [9]. Tensor $\mathbf{A}$ size of $n \times n \times n_3$ is called f-diagonal tensor if each frontal slice $\mathbf{A}^{(i)}$ is a diagonal matrix.

**Theorem 1** Tensor singular value decomposition (t-SVD) [9, 27]. The t-SVD of $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is given by

$$\mathbf{A} = \mathbf{U} \ast \mathbf{S} \ast \mathbf{V}^T,$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal tensors of size $n_1 \times n_1 \times n_3$ and $n_2 \times n_2 \times n_3$, respectively, $\mathbf{S}$ is a rectangular f-diagonal tensor of size $n_1 \times n_2 \times n_3$. The t-SVD for the order-3 tensor case is shown in Figure 1.

**Definition 6** Tensor multi-rank and tubal-rank [9]. The multi-rank of a tensor $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a vector $m \in \mathbb{N}^{n_1 \times 1}$ with its $i$th entry as the rank of the $i$th frontal slice in Fourier domain, that is $m_i = \text{rank}(\mathbf{A}^{(i)})$. The tensor tubal-rank $r(\mathbf{A})$, is defined as the number of non-zero singular tubes of $\mathbf{S}$, where $\mathbf{A} = \mathbf{U} \ast \mathbf{S} \ast \mathbf{V}^T$. The symbol $\#$ below describes the number of $i$ satisfying the condition $\mathbf{S}(i,i,:)$.

$$r(\mathbf{A}) = \{i : \mathbf{S}(i,i,:) \neq 0\} = \max_i m_i$$

In t-SVD, the main information features of tensors can be grasped by keeping only the non-zero singular tubes in $\mathbf{S}$, for example, by truncating t-SVD to achieve dimensionality reduction, and it can approximate the original tensor. According to Theorem 1, the reduced (truncated) t-SVD of $\mathbf{A}$ is given by

$$\mathbf{A}_r = \mathbf{U}_r \ast \mathbf{S}_r \ast \mathbf{V}_r^T,$$

where $\mathbf{U}_r$ and $\mathbf{V}_r$ are orthogonal tensors of size $n_1 \times r \times n_3$ and $n_2 \times r \times n_3$, respectively, $\mathbf{S}_r$ is a rectangular f-diagonal tensor of size $r \times r \times n_3$. If $r = \min\{n_1, n_2\}$, the third-order tensor $\mathbf{A}$ is of full tubal-rank.

**Figure 1** The t-SVD of an $n_1 \times n_2 \times n_3$ tensor.
Definition 7 The tensor nuclear norm (TNN) \[6, 9\]. TNN norm \(\|\mathcal{A}\|_{\text{TNN}}\) is defined as the sum of the singular values of all frontal slices of \(\mathcal{A}\), and is a convex relaxation of the tensor tubal-rank \[7, 8\].

\[\|\mathcal{A}\|_{\text{TNN}} = \|\text{blockdiag}(\mathcal{A})\|_* = \|\mathcal{A}\|. \tag{5}\]

Definition 8 Inner product of tensors \[6\]. The inner product between the order-3 tensors \(\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}\) and \(\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}\) is defined as:

\[\langle \mathcal{A}, \mathcal{B} \rangle = \frac{1}{n_3} \text{trace}(\mathcal{A}^H \mathcal{B}) = \frac{1}{n_3} \langle \mathcal{A}, \mathcal{B} \rangle,\]

where \(\frac{1}{n_3}\) comes from the normalisation constant of the FFT. Then, the Frobenius norm \(\|\mathcal{A}\|_F\) is induced as:

\[\langle \mathcal{A}, \mathcal{A} \rangle = \frac{1}{\sqrt{n_3}} \|\mathcal{A}\|_F.\]

The definition of tensor inner product makes it possible to simplify tensor analysis with matrix analysis techniques.

Remark 1 Tensor operator via t-product [9]. For a tensor operator denoted by a tensor \(\mathcal{C}\) with the size of \(n_4 \times n_1 \times n_3\), which means mapping tensor \(\mathcal{B}\) with the size of \(n_1 \times n_2 \times n_3\) to \(\mathcal{A}\) with the size of \(n_4 \times n_2 \times n_3\) via t-product, we can transform it into the equivalent form in Fourier domain for computational efficiency as follows,

\[\mathcal{A} = \mathcal{C} \ast \mathcal{B} = \mathcal{A} \ast \mathcal{B},\]

where \(\mathcal{A}\) is a matrix of size \(n_2 n_3 \times n_2 n_3\), \(\mathcal{B}\) is a matrix of size \(n_1 n_3 \times n_2 n_3\), and \(\mathcal{C}\) is a matrix of size \(n_1 n_2 \times n_1 n_3\).

Definition 9 Ortho-normal basis. The ortho-normal basis \(\{\mathcal{W}_a\}_{a=1}^{n_1 n_3 n_3}\) composed of the basis \(\mathcal{W}_a\) with size \(n_1 \times n_2 \times n_3\) meets the orthonormality, that is \(\|\mathcal{W}_a\|_F = 1\) for all \(a = 1, 2, \ldots, n_1 n_2 n_3\) and \(\mathcal{W}_i \circ \mathcal{W}_j = 0\), \(i \neq j\), where \(\circ\) denotes Hadamard product and \(i, j \in \{1, 2, \ldots, n_1 n_2 n_3\}\).

Definition 10 Standard operator basis [26]. The basis \(\{\mathcal{W}_a\}_{a=1}^{n_1 n_3 n_3}\) composed of the basis \(\mathcal{W}_a\) with size \(n_1 \times n_2 \times n_3\) is standard operator basis, if each basis \(\mathcal{W}_a\) has only non-zero element 1 at the corresponding \(a_{ib}\) location index of \(\mathcal{W}_a(\cdot)\). The standard operator basis meets the orthonormality, is a special case of ortho-normal basis.

3 | THEOREM AND METHOD

Given the fact that an unknown tensor \(\mathcal{X}\) can be expanded in terms of an ortho-normal basis \(\{\mathcal{W}_a\}_{a=1}^{n_1 n_2 n_3}\) (referred to as an arbitrary operator basis\(^\text{5}\)) below:

\[\mathcal{X} = \sum_{a=1}^{n_1 n_2 n_3} (\mathcal{W}_a, \mathcal{X})\mathcal{W}_a\]

where \(\langle \cdot, \cdot \rangle\) denotes tensor inner product defined in Definition 8. \(\mathcal{W}_a\) and \(\mathcal{X}\) have the same size of \(n_1 \times n_2 \times n_3\). The expansion coefficients of \(\mathcal{X}\) with respect to the basis \(\{\mathcal{W}_a\}_{a=1}^{n_1 n_2 n_3}\) is \(\langle \mathcal{W}_a, \mathcal{X} \rangle\), where \(a = 1, 2, \ldots, n_1 n_2 n_3\).

If the expansion coefficients are partially known, how can \(\mathcal{X}\) be reconstructed from them? Obviously, this is an ill-posed problem. Generally, regularizations are employed to enforce the prior knowledge of \(\mathcal{X}\) to solve this problem. Based on the observation that \(\mathcal{X}\) is a low-rank tensor, we propose to reconstruct \(\mathcal{X}\) by solving the following convex optimization problem:

\[
\min_{\mathcal{X}} \|\mathcal{X}\|_{\text{TNN}} \text{ subject to } \mathcal{X} \Omega = \mathcal{N} \Omega, \forall a \in \Omega.
\]

where \(\langle \mathcal{W}_a, \mathcal{N} \rangle\) is the partial known coefficients, \(\mathcal{X}\) is the unknown tensor to be recovered, both \(\mathcal{X}\) and \(\mathcal{N}\) are order-3 tensors with the size of \(n_1 \times n_2 \times n_3\). \(\Omega \subset [1, n_1 n_2 n_3]\) is a random index set of size \(m\), which means that the coefficients \(\langle \mathcal{W}_a, \mathcal{N} \rangle\) are known for all \(a \in \Omega\). Here, tensor nuclear norm (TNN) is used to enforce the tensor tubal-rank defined in t-SVD framework.

If \(\{\mathcal{W}_a\}_{a=1}^{n_1 n_2 n_3}\) is the standard operator basis, the coefficients are elements of the tensor itself. Paper [6] has given the answer to this case by solving the optimization problem below.

\[
\min_{\mathcal{X}} \|\mathcal{X}\|_{\text{TNN}} \text{ subject to } \mathcal{X} \Omega = \mathcal{N} \Omega.
\]

As shown in Figure 2, TNN norm \(\|\mathcal{X}\|_{\text{TNN}}\) is defined as the sum of the singular values of all frontal slices of \(\mathcal{X}\). We use block diagonal procedure for the calculation of TNN norm, which saves computation time compared with directly constraining the tubal rank. The procedure of block diagonal constraint is shown in Figure 2. The rank of block diagonal matrix satisfies: \(\text{rank}(\mathcal{X}) \leq \min(n_1 n_3, n_2 n_3)\). This concept runs through the procedure of proof. When tensor \(\mathcal{X}\) is rearranged into the tensor with the size of \(n_2 \times n_1 \times n_3\), we get a new matrix \(\mathcal{X}\), the rank of this new \(\mathcal{X}\) is the same as that of the previous \(\mathcal{X}\). But when tensor \(\mathcal{X}\) is rearranged into the tensor with the size of \(n_2 \times n_3 \times n_1\) or \(n_3 \times n_1 \times n_2\),

\[^{5}\text{A detailed description of Golfing scheme is provided in Appendix 7.3.}\]
n_2 \times n_1 \text{ or } n_3 \times n_1 \times n_2 \text{ or } n_1 \times n_3 \times n_2 \text{, the rank of the new } \mathcal{X} \text{ is not the same as that of the previous } \mathcal{X}. \text{ This is also the reason why we use } \langle n, n, n \rangle \text{ to simplify the representation of } \langle n_1, n_2, n_3 \rangle \text{ in some places, where } n = \min(n_1, n_2). 

In order to recover } \mathcal{N} \text{ from the partially known coefficients by solving the problem of Equation (6), the question needs be addressed is: given that the tensor tubal-rank } r \leq n \text{ and } n = \min(n_1, n_2), \text{ how many randomly chosen coefficients are needed to efficiently reconstruct tensor } \mathcal{X} \text{ of size } n_1 \times n_2 \times n_3? 

Based on previous conclusions from the low-rank matrix completion [30, 31], it is easy to deduce that tensor } \mathcal{X} \text{ cannot be reconstructed if } \mathcal{X} \text{ has very few non-zero coefficients with respect to the basis } \{ \mathcal{W}_a \}. \text{ To ensure that each coefficient contains enough non-trivial information, some conditions are needed to characterise the incoherence between the tensor and the basis. Intuitively, tensors with small operator norm are 'incoherent' to all low-tubal rank tensors simultaneously. The 'incoherence conditions' have been proposed in tensor completion using t-SVD [6]. The 'incoherent' has a well-known analogue in compressed sensing [32–34]. There, one uses the fact that 'vectors with small entries' are incoherent to 'sparse vectors'. Ref. [26] proved that matrices with small operator norm are 'incoherent' to all low rank matrices simultaneously. The definition of coherence stated below is closely related to, but more general than, the parameter } \mu_0 \text{ used in ref. [6]. We follow the form of definitions in ref. [20] and define coherence conditions of a tensor, which are extensions of coherence condition of a matrix. Then, we give the answer to the above question under the defined tensor coherence conditions. 

To state our results clearly, we need to introduce some notation. We introduce the orthogonal decomposition } \mathbb{R}^{n_1 \times n_2 \times n_3} = T \oplus T^\perp \text{ where } T \text{ is the linear space spanned by the elements of the form } U(:,;,:,;,:) \ast \mathcal{X}^\perp \text{ and } V \ast \mathcal{Y}(:,;,:,;,:)^\perp. \text{ } \mathcal{X}, \mathcal{Y} \text{ are arbitrary tensor columns with the size of } n_1 \times 1 \times n_3, n_2 \times 1 \times n_3 \text{, respectively, and } k = 1, 2, \ldots, r. \text{ The orthogonal projection } \mathcal{P}_T \text{ onto } T \text{ is given as follows,} 

\begin{align*}
\mathcal{P}_T(A) &= U \ast U^\perp \ast A + A \ast V \ast V^\perp - U \ast U^\perp \ast A \ast V \ast V^\perp 
\end{align*}

and the projections onto the orthogonal complement } T^\perp \text{ is given as follows,}

\begin{align*}
\mathcal{P}_{T^\perp}(A) &= A - \mathcal{P}_T(A) \\
&= (I - U \ast U^\perp) \ast A \ast (I - V \ast V^\perp)
\end{align*}

\text{FIGURE 2 The procedure of block diagonal constraint}

\begin{equation}
P_{T^\perp}A = A - \mathcal{P}_T(A) = (I - U \ast U^\perp) \ast A \ast (I - V \ast V^\perp) 
\end{equation}

Define a random variable } \delta_a = 1_{a \in \Omega} \text{ where } 1_{(\cdot)} \text{ is the indicator function. Let } \mathcal{R}_\Omega: \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^{n_1 \times n_2 \times n_3} \text{ be a random projection as follows,}

\begin{equation}
\mathcal{R}_\Omega(A) = \sum_{a = 1}^{n_1 n_2 n_3} \frac{1}{\rho} \delta_a(\mathcal{W}_a, A)\mathcal{W}_a 
\end{equation}

where } \rho = \frac{m}{n} \text{ is the probability of that } a \in [1, n_1 n_2 n_3] \text{ is included in } \Omega. \text{ Then, we have } \mathcal{P}_{T^\perp}(A) = \sum_{a = 1}^{n_1 n_2 n_3} \delta_a(\mathcal{W}_a, A)\mathcal{W}_a. 

As in the previous work [6, 30, 31], the sampling model we use is the Bernoulli model. As we know, there are also other commonly used random models, like sampling with replacement and sampling without replacement. Similar to matrix completion problems, researchers can get corresponding recovery guarantees by slightly changing the proof procedure [31, 35]. 

Next, we present our definition of coherence conditions and new theorem for tensor reconstruction problem (Equation 6) using tensor-singular value decomposition (t-SVD).

**Definition 11** Tensor coherence conditions. The order-3 tensor } \mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3} \text{ has coherence } } \nu \text{ with respect to an operator basis } \{ \mathcal{W}_a \}_{a = 1}^{n_1 n_2 n_3} \text{ if either}

\begin{equation}
\max_a \| \mathcal{W}_a \|_2^2 \leq \nu \frac{1}{n} 
\end{equation}

or the two estimates

\begin{equation}
\max_a \| \mathcal{P}_T \mathcal{W}_a \|_F^2 \leq 2\nu \frac{r}{n}, 
\end{equation}

\begin{equation}
\max_a (\mathcal{W}_a, U \ast V^\perp)^2 \leq \nu \frac{r}{n^2 n_3} 
\end{equation}

hold, where } n = \min(n_1, n_2) \text{ and } U \ast S \ast V^\perp \text{ is the reduced t-SVD of } \mathcal{M}. 

The first inequality in Definition 11 is the tensor coherence condition for reconstructing low tubal-rank tensors from limited Fourier-type coefficients. Inequalities (Equations 12, 13) are suitable for the general case that is reconstructing low
tubal-rank tensors from the limited coefficients with respect to any ortho-normal basis.

**Theorem 2** Suppose \( M \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) and its reduced t-SVD is given by \( M = U \ast S \ast V \) where \( U \in \mathbb{R}^{n_1 \times r} \), \( S \in \mathbb{R}^{r \times r} \), and \( V \in \mathbb{R}^{n_3 \times r} \). Suppose \( M \) satisfies the coherence condition above with parameter \( \nu > 0 \). Then there exist constants \( c_0, c_1, c_2 > 0 \) such that if

\[
p \geq c_0 \nu \log^2(n_3(n_1 + n_2)) / \min\{n_1, n_2\}
\]  

(14)

then \( M \) is the unique minimizer to Equation (6) with probability at least \( 1 - c_1((n_1 + n_2)n_3)^{-c_2} \). Theorem 2 is the main theoretical result.

Reference [6] concentrated only on the problem of tensor completion where it aims to recover a low-tubal rank tensor from randomly selected tensor elements. While our paper extends tensor completion to a more general scenario where the coefficients under any ortho-normalised basis are randomly sampled to recover the tensor. The tensor recovery problem we focus on can be seen as a generalisation of the tensor completion problem in ref. [6]. Similar to ref. [6], we also have to handle the block diagonal constraint in the proof procedure, which leads to subtle differences with matrix reconstruction problems in ref. [26]. We adopted the Golting scheme 3 in our proof which is widely used after [26] first introduced. The difference is in our generalisation, the critical step is how to verify that the condition (2, b) in Proposition 1 below still holds under any ortho-normalised basis. In addition, in our proof, we use the conclusion in Lemma 3 and mathematical induction to get Equation (36) in Appendix 7.2.

The algorithm is designed for solving Equation (6). We turn the constraints \( (X, W_a) = (N, W_a) \) into \( R \circ \psi X = B \), where \( \forall a \in \Omega \) and \( \circ \) denotes Hadamard product and \( R \) is the binary tensor contains 0 and 1. The index set of non-zero elements in \( R \) is \( \Omega \). \( \psi \) is an operator, which acts on \( X \) to get the expansion coefficients of \( X \) with respect to basis \( \{W_a\} \), for all \( a = 1, 2, \ldots, n_1 n_2 n_3 \). \( B \) is the undersampled coefficients. Then, we rewrite Equation (6) as an unconstrained convex optimization problem:

\[
\min_{X} \frac{1}{2} \| R \circ \psi X - B \|_F^2 + \lambda \| X \|_{TNN}
\]  

(15)

where \( \lambda \) is a regularisation parameter.

We develop an alternating direction method of multipliers (ADMM) algorithm to solve the problem Equation (15). We introduce an auxiliary variable \( Z = X \), the augmented Lagrangian function of Equation (15) is given by

\[
L(X, Z, Q) = \frac{1}{2} \| R \circ \psi X - B \|_F^2 + \lambda \| X \|_{TNN} + \frac{\rho}{2} \| X - Z \|_F^2 + \langle Q - X, X - Z \rangle
\]  

(16)

where \( Q \) denotes the Lagrangian multiplier, and \( \rho > 0 \) is called the penalty parameter. By applying ADMM, each sub-problem is performed at each iteration as follows:

\[
X_{t+1} = \arg \min_X \frac{1}{2} \| R \circ \psi X - B \|_F^2 + \frac{\lambda \rho}{2} \| X - Z_t \|_F^2 + \frac{1}{\rho} \| Q_t \|_F^2
\]  

(17)

\[
Z_{t+1} = \arg \min_Z \| Z \|_{TNN} + \frac{\rho}{2} \| X_{t+1} - Z \|_F^2 + \frac{1}{\rho} \| Q_t \|_F^2
\]  

(18)

\[
Q_{t+1} = Q_t + \rho (X_{t+1} - Z_{t+1})
\]  

(19)

The variables \( Z \) and \( Q \) are initialized as tensors fully filled with zeros before the above sub-problems can be alternately solved. In sub-problem (Equation 17) (linear least-squares problem), the value of \( X \) at the \((t + 1)_{th} \) iteration step is \( X_{t+1} \), which can be obtained by taking derivative of Equation (16) with respect to \( X \) and making this derivative equal to zeros:

\[
X_{t+1} = \psi'(B + \lambda \psi(\rho Z_t - Q_t)) \circ (R + \lambda \rho)^{-1}
\]  

(20)

where \( \psi' \) is the inverse operator of \( \psi \), \( \psi \psi = I \), and \( I \) denotes the tensor with every entry being 1.

The details of solving sub-problem (Equation 18) are as follows. It is well known that block circulant matrices can be block diagonalized by using the Fourier transform. This property can be used to simplify the calculation of t-SVD [9, 27, 28]. We compute t-SVD decomposition using matrix SVDs in Fourier domain and give the algorithm for solving the problem (Equation 18) as shown in Algorithm 1, where shrink(S) is singular value thresholding (SVT) function [36, 37].

**Algorithm 1** For solving problem Equation (18)

**Input:** \( M = X_t + \frac{1}{\rho} Q_t, \frac{1}{\rho} \)

**Output:** \( Z_{t+1} \)

1: \( D \leftarrow \text{fft}(M, [ ], 3) \)
2: for \( i = 1 \) to \( n_3 \) do
3: \( [U^{(i)} S, V] = \text{svd}(D^{(i)}) \)
4: \( \hat{U}^i = U; \hat{S}^{(i)} = \text{shrink}_\lambda(S); \hat{V}^{(i)} = V \)
5: end for
6: \( U \leftarrow \text{iff}(\hat{U}, [ ], 3); S = \text{iff}(\hat{S}, [ ], 3); V = \text{iff}(\hat{V}, [ ], 3) \)
7: \( Z_{t+1} = U \ast S \ast V^T \)
The whole developed ADMM algorithm named as t-SVD method for solving the problem (Equation 15) is shown in Algorithm 2.

Algorithm 2 t-SVD method

**Input:** Randomly measurements $\mathcal{E}$. The sampling mask $\mathcal{R}$, maximum number of iteration $t_{\text{max}}$, convergence condition $\eta_{\text{tol}}$

1. Parameters: $\lambda$, $\rho$.
2. While $t \leq t_{\text{max}}$ or $\eta_t < \eta_{\text{tol}}$ do:
   3. Solve Equation (17) for $\lambda_{t+1}$.
   4. Solve Equation (18) for $Z_{t+1}$.
   5. Update $Q_{t+1}$ by Equation (19).
6. $\eta_{t+1} = \frac{\|X_{t+1} - X_t\|_F}{\|X_t\|_F}$
7. End

**Output:** $\lambda^*$, here * represents the optimal solution.

4 | PROOF OF THEOREM 2

In order to prove Theorem 2, we need to prove that $\|\mathcal{M} + Z\|_{\text{TN}} > \|\mathcal{M}\|_{\text{TN}}$, for any $Z$ supported in $\Omega'$. Since $\|\mathcal{M} + Z\|_{\text{TN}} > \|\mathcal{M}\|_{\text{TN}}$, for any $\lambda \neq \lambda$ obeying $\mathcal{P}_{\Omega}(\lambda - \mathcal{M}) = 0$, we have $\|\lambda\|_{\text{TN}} > \|\mathcal{M}\|_{\text{TN}}$, which proves $\mathcal{M}$ is the unique minimizer of Equation (6). Based on Proposition 1 and Lemma 1 below, we prove $\|\mathcal{M} + Z\|_{\text{TN}} > \|\mathcal{M}\|_{\text{TN}}$ in Appendix 7.1. Since Proposition 1 and Lemma 1 support the main proof of $\|\mathcal{M} + Z\|_{\text{TN}} > \|\mathcal{M}\|_{\text{TN}}$, we first prove Proposition 1 and Lemma 1 in this section and Appendix 7.2.

And as proved in this section and Appendix 7.2, for some constants $c_1 > 0$ and $c_2 > 0$, if $\rho$ satisfies Equation (14), the conditions (1) and (2) in Proposition 1 are satisfied with probability at least $1 - c_1 ((n_1 + n_2)\sigma_3)^{-c_2}$. So, we drive that the $\mathcal{M}$ in Theorem 2 would be the unique minimizer to Equation (6) with probability at least $1 - c_1 ((n_1 + n_2)\sigma_3)^{-c_2}$. Then, the proof of our Theorem 2 is finished.

**Proposition 1** Suppose $\rho$ satisfies Equation (14), then tensor $\mathcal{M}$ is the unique minimizer to Equation (6) if the following conditions hold:

1. $\|\mathcal{M} + Z\|_{\text{TN}} > \|\mathcal{M}\|_{\text{TN}}$
2. There exists a dual certificate tensor $\mathcal{Y}$, $\mathcal{P}_{\Omega}(\mathcal{Y}) = \mathcal{Y}$ and a) $\|\mathcal{P}_{\Omega}(\mathcal{Y}) - U + V^T\|_F \leq \frac{1}{4n_1}$
   b) $\|\mathcal{P}_{\Omega^c}(\mathcal{Y})\| \leq \frac{1}{2}$

**Lemma 1** Suppose $\|\mathcal{P}_{\Omega} \mathcal{R} G \mathcal{P}_{\mathcal{Y}} - \mathcal{P}_{\mathcal{Y}}\|_F \leq \frac{1}{2}$, then for any tensor $Z$ and $\mathcal{P}_{\Omega}(Z) = 0$, we have

$$\frac{1}{2} \|\mathcal{P}_{\Omega^c}(Z)\|_{\text{TN}} > \frac{1}{4n_1} \|\mathcal{P}_{\mathcal{Y}}(Z)\|_F$$

For the proof of Proposition 1 and Lemma 1, we first give the proof for Fourier-type basis in this section, then present the relatively simple modifications to cover the general case in Appendix 7.2.

As we all know, Fourier basis is a typical vector basis with small vector spectral norm which satisfies the incoherent condition. In ref. [26], the matrix basis is called Fourier-type basis as long as the operator norm satisfies relatively small conditions. Similarly, we define the tensor basis of the condition in the operator norm satisfying Equation (11) as the Fourier-type basis. First, we prove that Theorem 2 holds in the case of Fourier-type basis. Then, we modify the proof in two places and propose two new lemmas (Lemma 3 and Lemma 4) to prove that Theorem 2 holds in the case of any ortho-normal basis.

As mentioned above, the inequality (Equation 11) in Definition 11 is the coherence condition for reconstructing low tubal rank tensors from limited Fourier-type coefficients. Next, we show that the construction of $\mathcal{Y}$ using Golging scheme continues to work if assumption (Equation 11) on the operator norm of the basis’ elements.

Before continuing, the theorem below will be used in the proof of Proposition 1(1) and Lemma 1–4, which was first developed in ref. [38].

**Theorem 3** [38] Let $X_1, X_2, \ldots, X_L$ be independent zero-mean random matrices of dimension $d_1 \times d_2$.

**Suppose**

$$\rho^2 = \max \left\{ \left\| \mathcal{E} X_k X_k^T \right\|, \left\| \mathcal{E} X_k^T X_k \right\| \right\}$$

and

$$\|X_k\| \leq M$$

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

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

(22)

Theorem 3 is a corollary of Chernoff bound for finite dimension operators developed from ref. [39]. Usually Equation (22) is called Non-commutative Bernstein Inequality (NBI). Reference [40] gave an extension of Theorem 3, stated that if

$$\max \left\{ \left\| \mathcal{E} \sum_{k=1}^L X_k X_k^T \right\|, \left\| \mathcal{E} \sum_{k=1}^L X_k^T X_k \right\| \right\} \leq \sigma^2$$

and let
Proof: As in the previous work [6, 30, 31], we employ the Bernoulli model as the sampling model. Then $\mathbb{E} R_{\Omega} = I$. Note that

$$\|T_a - \frac{1}{n^2 n_3} P_T\|_{op} = \|T_a - \frac{1}{n^2 n_3} P_T\|_{op} \leq \frac{\text{max}\{\frac{1}{p} \|P_T W_a\|_{F}, \frac{1}{n^2 n_3}\}}{n^2 n_3} \leq \frac{2 \nu r}{np}$$

We derive the above inequality by using the coherence assumption (Equation 12) and the fact that if $A$ and $B$ are positive semi-definite matrices, then $\|A - B\| \leq \max\{\|A\|, \|B\|\}$. From Equation (25) we have $\mathbb{E} T_a = \frac{1}{n^2 n_3} P_T$. So

$$\mathbb{E}\left[\left(T_a - \frac{1}{n^2 n_3} P_T\right)\left(T_a - \frac{1}{n^2 n_3} P_T\right)^\top\right] = \mathbb{E}\left[\left(T_a - \frac{1}{n^2 n_3} P_T\right)^2\right] = \mathbb{E}\left[\left(T_a - \frac{1}{n^2 n_3} P_T\right)^2\right]_{op}$$

Given any tensor $Z$ of size $n \times n \times n_3$, we can decompose $P_T(Z)$ as the following

$$P_T(Z) = \sum_{a=1}^{n^2 n_3} \langle P_T(Z), W_a \rangle W_a$$

This gives

$$R_{\Omega} P_T(Z) = \sum_{a=1}^{n^2 n_3} \frac{1}{p} \delta_a(Z, P_T W_a) W_a$$

and

$$P_T R_{\Omega} P_T(Z) = \sum_{a=1}^{n^2 n_3} \frac{1}{p} \delta_a(Z, P_T W_a) P_T W_a$$

which implies

$$P_T R_{\Omega} P_T(Z) = \sum_{a=1}^{n^2 n_3} \frac{1}{p} \delta_a(Z, P_T W_a) P_T W_a$$

Define operator $T_a$ which maps $Z$ to $\frac{1}{p} \delta_a(Z, P_T W_a) P_T W_a$, then observe that $\|T_a\|_{op} = \|T_a\| \leq \frac{\text{max}\{\frac{1}{p} \|P_T W_a\|_{F}, \frac{1}{n^2 n_3}\}}{n^2 n_3} \leq \frac{2 \nu r}{np}$, and $\|P_T\|_{op} = \|P_T\| \leq 1$, we have

$$\mathbb{P}\left[\|P_T R_{\Omega} P_T - P_T\|_{op} > \tau\right] = \mathbb{P}\left[\sum_{a=1}^{n^2 n_3} \left(T_a - \frac{1}{n^2 n_3} P_T\right) \| > \tau\right] \leq 2 n n_3 \exp\left(\frac{7 \beta \nu r \log^2(2n n_3)}{3} \left(\frac{np}{2 \nu r} + \frac{np}{2 \nu r} + \frac{6}{np}\right)\right) \leq (2n n_3)^{1-2\theta}$$

which finishes the proof.
4.2 Proof of Lemma 1

Proof: Given any tensor $Z$ as a perturbation with $\mathcal{P}_\Omega(Z) = 0$ and according to Proposition 1(1): $\|P_T \mathcal{R}_\Omega P_T - P_T\|_\infty \leq \frac{1}{n}$, we have

$$\|\sqrt{p} R_\Omega P_T(Z)\|_F^2 = \frac{1}{n^3} \|\sqrt{p} R_\Omega P_T(Z)\|_F^2$$

$$= \frac{1}{n^3} \langle R_\Omega P_T(Z), p R_\Omega P_T(Z) \rangle$$

$$= \frac{1}{n^3} \langle P_T R_\Omega P_T(Z), Z \rangle$$

$$= \frac{1}{n^3} \left( \langle P_T R_\Omega P_T - P_T \rangle Z, P_T(Z) \rangle + \langle P_T(Z), P_T(Z) \rangle \right)$$

$$\geq \frac{1}{n^3} \left( \|P_T(Z)\|_F^2 - \|P_T R_\Omega P_T - P_T\| \|P_T(Z)\|_F^2 \right)$$

$$\geq \frac{1}{2n^3} \|P_T(Z)\|_F^2 = \frac{1}{2} \|P_T(Z)\|_F^2$$

Combining the last two display equations gives

$$\frac{1}{p} \|P_T(Z)\|_F^2 \geq \frac{1}{2} \|P_T(Z)\|_F^2$$

and since

$$\|P_T Z\|_{\text{TNN}} = \|P_T Z\|_F \geq \|P_T Z\|_F = \sqrt{n^3}$$

we get $\frac{1}{4} \|P_T Z\|_{\text{TNN}} \geq \frac{1}{2} \|P_T(Z)\|_F$. So, under the conditions of Theorem 2, $p \geq \frac{1}{2(2nm)^2}$ holds clearly, which finish the proof.

4.3 Proof of Proposition 1(2)

Before validating Proposition 1(2), we need to introduce Lemma 2 which will be used in this part.

**Lemma 2** If $p$ satisfies Equation (14) in Theorem 2, and $Z \in T$ sized $n \times n \times n_3$. Then for some constant $c_3$, we have

$$\|P_T \mathcal{R}_\Omega Z\|_F \leq \frac{1}{4\sqrt{r}} \|Z\|_F$$

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

Proof of Proposition 1(2):

a. Here, we construct a dual certificate $\mathcal{Y}$ and show it satisfies both conditions, that is, Proposition 1(2). We use the approach called Golfing scheme introduced in ref. [26] and construct the tensor dual certificate $\mathcal{Y}$ iteratively follows the idea in refs. [6, 41]. A more detailed description of Golfing scheme is provided in Appendix 7.3.

Let $\Omega$ be a union of smaller sets $\Omega_i$ such that $\Omega = \bigcup_{i=1}^{\nu} \Omega_i$, where $t_0 := 2 \log(2nm)$. For each $t$, we assume

$$\mathbb{P}[a \in \Omega_t] = q := 1 - (1 - (1/4)^t)$$

and is independent of all others. Clearly, this $\Omega_t$ is equivalent to the original $\Omega$ in our Bernoulli model.

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

$$\mathcal{G}_t = \mathcal{G}_{t-1} + \mathcal{R}_\Omega(U \ast V^T - P_T(\mathcal{G}_{t-1}))$$

then set a tensor $\mathcal{Y} = \mathcal{G}_{t_0}$ and we have $\mathcal{P}_\Omega(\mathcal{Y}) = \mathcal{Y}$ by the above construction. Set $D_t = U \ast V^T - P_T(\mathcal{G}_t)$ for $t = 1, 2, \ldots, t_0$. By definition of $\mathcal{G}_t$, we have $D_0 = U \ast V^T$ and by the deductive method we have

$$D_t = (P_T - P_T \mathcal{R}_\Omega P_T)(D_{t-1})$$

Note that $\Omega_t$ is independent of $D_t$, and if $p$ satisfies Equation (14) in Theorem 2, then $q \geq p/t_0 \geq c_3 \sqrt{\nu \log(2nm^3)/n}$. According to Proposition 1(1), we have

$$\|D_t\|_F \leq \|P_T - P_T \mathcal{R}_\Omega P_T\| \|D_{t-1}\|_F \leq \frac{1}{2} \|D_{t-1}\|_F$$

for each $t$. Applying the above inequality recursively, we get

$$\|D_0\|_F \leq \left(\frac{1}{2}\right)^{t_0} \|D_0\|_F$$. Then, according to the property

$$\|U \ast V^T\|_F \leq \sqrt{r}$$

gives

$$\|P_T(\mathcal{Y}) - U \ast V^T\|_F \leq \|D_0\|_F$$

$$\leq \left(\frac{1}{2}\right)^{t_0} \|U \ast V^T\|_F \leq \frac{1}{4nm^3} \sqrt{r} \leq \frac{1}{4nm^3}$$

holds with probability at least $1 - c'(2nm)^{-c'}$ by the union bound, for some positive constants $c', c''$ large enough.
From Equation (27) we know that
\[ \mathcal{Y} = \mathcal{G}_m = \sum_{t=1}^{t_0} \mathcal{R}_\Omega \mathcal{D}_{t-1}, \]
so use Lemma 2 we obtain
\[ \| P_{T^t} (\mathcal{Y}) \|_F \leq \sum_{t=1}^{t_0} \| P_{T^t} \mathcal{R}_\Omega \mathcal{D}_{t-1} \|_F \leq \frac{1}{4\sqrt{r}} \sum_{t=1}^{t_0} \| D_{t-1} \|_F \]
\[ \leq \frac{1}{4\sqrt{r}} \sum_{t=1}^{t_0} \left( \frac{1}{2} \right)^{t-1} \| D_0 \|_F \leq \frac{1}{2} \]
holds with probability at least \( 1 - c_1 (2n_3)^{-c_2} \) by union bound for some large enough constants \( c_1, c_2 > 0 \).

### 4.4 Proof of Lemma 2

Observe that
\[ P_{T^t} \mathcal{R}_\Omega \mathcal{Z} = P_{T^t} \left( \sum_a \frac{1}{p} \mathcal{R}_\Omega \mathcal{D}_a \right) \]
\[ = \sum_a \frac{1}{p} \delta_a \mathcal{D}_a P_{T^t} \mathcal{W}_a \]
define
\[ \mathcal{H}_a (\mathcal{Z}) = \frac{1}{p} \delta_a \mathcal{D}_a P_{T^t} \mathcal{W}_a \]
then we get \( P_{T^t} \mathcal{R}_\Omega \mathcal{Z} = \sum_a \mathcal{H}_a \mathcal{Z} \). We use the Bernoulli model.

So, \( E[\delta_a] = E[\delta_a^2] = p \), and \( E[\mathcal{R}_\Omega] = \mathcal{I} \). Note that if \( \mathcal{Z} \in T \), then \( P_{T^t} \mathcal{Z} = 0 \),

\[ E[\mathcal{H}_a (\mathcal{Z})] = \frac{1}{n^2 n_3} E[P_{T^t} \mathcal{R}_\Omega \mathcal{Z}] = \frac{1}{n^2 n_3} P_{T^t} \mathcal{Z} = 0 \]
and use the fact that \( \| P_{T^t} \mathcal{W}_a \| \leq \| \mathcal{W}_a \| \) (since \( \mathcal{W}_a = P_{T^t} \mathcal{W}_a + P_{T^t} \mathcal{W}_a )\), we have

\[ \| \mathcal{H}_a (\mathcal{Z}) \| = \| \mathcal{H}_a (\mathcal{Z}) \| = \left( \frac{1}{p} \delta_a \mathcal{D}_a P_{T^t} \mathcal{W}_a \right) \]
\[ \leq \frac{1}{p} \| \mathcal{D}_a \| \| \mathcal{W}_a \| \| P_{T^t} \mathcal{W}_a \| \]
\[ \leq \frac{1}{n^2 n_3} \| \mathcal{D}_a \| \| \mathcal{W}_a \| \| P_{T^t} \mathcal{W}_a \| \]
\[ \leq \frac{1}{n^2 n_3} \| \mathcal{D}_a \| \| \mathcal{W}_a \| \| P_{T^t} \mathcal{W}_a \| \]
\[ \leq \frac{1}{n^2 n_3} \| \mathcal{W}_a \|^2 \| \mathcal{Z} \|_F \| P_{T^t} \mathcal{W}_a \| \]
\[ \leq \frac{1}{n^2 n_3} \| \mathcal{W}_a \|^2 \sqrt{2n_3 r} \| \mathcal{Z} \|_F = \left( \frac{1}{p} \right) \| \mathcal{W}_a \|^2 \sqrt{2r} \| \mathcal{Z} \|_F \]
\[ \leq \frac{2r \sqrt{r}}{n} \| \mathcal{Z} \|_F \]

At the end of the above inequalities, we use the inequality Equation (11). Using Equation (11), we also have

![Figure 3](image-url)

**Figure 3** Reconstruction of order-3 tensor \((32 \times 32 \times 20)\) with different tubal rank \(r\) from their coefficients in db4 basis with different sampling rates. In the figures on the left, the white cell stands for exact reconstruction, and black one stands for the failure. The figures on the right depict the RLNE curves of one typical run of the simulation. The value of each cell is the RLNE of the recovery under the corresponding sampling rate and tubal rank. The colour scale ranges from 0 to 0.55.
\[
\| \mathbb{E} \left[ \sum_a \mathcal{H}^2_a \right] \| = \| \mathbb{E} \left[ \sum_a \mathcal{H}^2_a \right] \|
\]
\[
= \| \mathbb{E} \left[ \frac{1}{p} \delta^2 \sum_a (W_a, \mathcal{Z})^2 (P_T \mathcal{W}_a)^2 \right] \|
\]
\[
\leq \frac{1}{p} \| \mathcal{Z} \|^2 \| W_a \|^2 \leq \frac{1}{p} \nu \| \mathcal{Z} \|^2
\]
then use Theorem 3, we have

\[
\| P_T^* R_{\mathcal{A}} \mathcal{Z} \|_\infty \leq \left( \sqrt{\frac{\nu \log(2nm_3)}{pn}} + c \frac{\nu \log(2nm_3)}{pn} \right) \| \mathcal{Z} \|_F
\]

holds with probability at least \( 1 - (2nm_3)^{-(c-1)} \) for any constant \( c \), given \( p \) satisfying Equation (14), which finish the proof of Lemma 2.
5 | EXPERIMENTAL RESULTS

We conduct experiments on random tensors and dynamic MR images to estimate the performance of our methods. All simulations were carried out on Windows 10 and MATLAB R2016a running on a PC with an Intel Core i5 CPU 3.2 GHz and 12 GB of memory. For quantitative evaluation, the reconstruction quality was measured by the relative least normalised error (RLNE) and structural similarity (SSIM).

RLNE is a standard image quality metric indicating the difference between the reconstruction $\tilde{\mathcal{X}}$ and the original tensor $\mathcal{X}$:

$$
\zeta_{RLNE} = \frac{\|\mathcal{X} - \tilde{\mathcal{X}}\|_F}{\|\mathcal{X}\|_F}
$$

The structural similarity (SSIM) index\(^4\) [42] attempts to measure the change in luminance, contrast, and structure between $\tilde{\mathcal{X}}(\cdot, \cdot, i)$ and $\mathcal{X}(\cdot, \cdot, i)$:

$$
\zeta_{SSIM} = \sum_{ii} \frac{(2\mu_{1,ii} + \mu_{2,ii} + \epsilon_1)(2\sigma_{12,ii} + \epsilon_2)}{(\mu_{1,ii}^2 + \mu_{2,ii}^2 + \epsilon_1)(\sigma_{1,ii}^2 + \sigma_{2,ii}^2 + \epsilon_2)}
$$

where, $i = 1, 2, \ldots, N_3$, $\mu_{1,ii}$ and $\mu_{2,ii}$ are mean intensities at the $ii$th local window of $\tilde{\mathcal{X}}(\cdot, \cdot, i)$ and $\mathcal{X}(\cdot, \cdot, i)$, while $\sigma_{1,ii}$ and $\sigma_{2,ii}$ are the corresponding SDs. $\sigma_{12,ii}$ denotes the covariance and the constants $\epsilon_1$, $\epsilon_2$ are included to avoid instability. Then, we use the mean SSIM $\bar{\zeta}_{SSIM} = \frac{1}{N_3} \sum_i \zeta_{SSIM}$ as a metric for the quality of the reconstructed tensor.

5.1 | Random tensor reconstruction

In order to evaluate the performance of our t-SVD methods, we conduct numerical simulations on real-valued random order-3 tensors with tubal rank exactly $r$. To create such a tensor, we generate an order-3 tensor with a.i.i.d. Gaussian distribution first, and then find its tubes via t-SVD. We keep the $r$ largest tubes but set the rest to zeros, then get the tensor $\mathcal{X}$ according to Equation (4). In the signal recovery application, the Fourier base is the most commonly used, while the application under other orthogonal bases is seldom used. We only show the experimental results under the orthogonal Fourier base and the db4 wavelet basis. Data acquisition is simulated by undersampling Fourier coefficients and db4 wavelet coefficients of the frontal slices $\mathcal{X}(\cdot, \cdot, k)$. The number of the measurements is quantified in terms of the percentage of the number of fully sampled coefficients, referred to as sampling rate (SR). Here, we use variable sampling masks with different sampling rates. We reconstruct these random tensors using t-SVD method and compute the relative least normalised error (RLNE). If the RLNE $\leq 0.05$, we claim the reconstruction is exact. The experimental results under the Fourier transform and db4 wavelet transform are shown in Figures 3 and 4, respectively. In the figures on the left, the white cell stands for exact reconstruction, and black one stands for the failure. The figures on the right depict the RLNE curves of one typical run of the simulation. The value of each cell is the RLNE of the recovery under the corresponding sampling rate and tubal rank. The colour scale ranges from 0 to 0.55.

As shown in Figures 3 and 4, more accurate results can be obtained from a widely range of lower tubal rank and lower sampling rate, while more accurate results can also be obtained from a widely range of higher tubal rank and higher sampling rate. Under the same tubal rank and sampling rate, the RLNE under orthogonal Fourier basis is smaller than that of db4 wavelet, which means that the orthogonal Fourier basis is better than that of db4 wavelet in the case studies.

The performance of the proposed t-SVD method depends on the regularisation parameter pair $(\lambda, \rho)$. We first generate a random order-3 tensor of size $30 \times 30 \times 20$ with tubal rank $r = 5$. Then, the parameters $(\lambda, \rho)$ were optimised based on this tensor using 1/2 undersampling variable density mask. We set the maximum number of iteration $t_{\text{max}} = 200$ and convergence condition $\eta_{\text{tol}} = 10^{-5}$. Under the orthogonal Fourier transform base, Figure 5 shows the parameter optimization based on RLNE from different pairs of $\lambda$ and $\rho$. It can be seen that the optimization setting for $\lambda$ is less than 0.1 while the low-rank regularisation setting $\rho$ is from 0.1 to 50.

5.2 | Image reconstruction

Based on Definition 6, we can show the low tubal rankness of the d-MRI data of size $n_1 \times n_2 \times n_3$ by plotting $\delta_i$ which is defined as follows.

$$
\delta_i = \frac{1}{n_3} \sum_{j=1}^{n_3} S(i, i, j), \quad i = 1, 2, \ldots, \min\{n_1, n_2\}
$$

[1] http://www.ece.uwaterloo.ca/~y70wang/research/ssim/
**FIGURE 6** Two d-MRI data sets and their low tubal rankness. The figures on the left show one frame of each data set and the figures on the right are the $\delta_i$ values.

**FIGURE 7** Sampling mask and original images. (a) The variable density sampling mask; (b) The 10th spatial frame of the gold standard Cardiac MR images 1; (c) The 136th time profile of the gold standard Cardiac MR images 1; (d) The 10th spatial frame of the gold standard Cardiac MR images 2; The gold standard images have been normalised such that the greyscale ranges from 0 to 1.

(a) Cardiac MR images 1 with the size of $256 \times 256 \times 25$, complex-valued.

(b) Cardiac MR images 2 with the size of $200 \times 256 \times 256$, real-valued.
TABLE 2 The RLNEs(%)/SSIMs of four methods with different sampling rates

| Methods     | Cardiac MR images 1 (%) | Cardiac MR images 2 (%) |
|-------------|-------------------------|-------------------------|
|             | 20%                     | 40%                     | 60%                     |
| Zero-filled | 36.79/0.7329            | 7.41/0.9761             | 2.38/0.9951             |
| Tucker      | 11.48/0.9538            | 4.89/0.9987             | 2.69/0.9962             |
| Unfold      | 6.22/0.9841             | 3.59/0.9933             | 2.01/0.9978             |
| t-SVD       | 5.63/0.9854             | 3.59/0.9932             | 2.13/0.9975             |
|             | 20%                     | 40%                     | 60%                     |
|             | 31.40/0.6589            | 8.87/0.9140             | 5.41/0.9642             |
|             | 15.30/0.8578            | 5.37/0.9653             | 3.27/0.9853             |
|             | 4.42/0.9789             | 2.01/0.9929             | 1.21/0.9972             |
|             | 4.65/0.9771             | 2.27/0.9916             | 1.36/0.9965             |

F I G U R E 8 RLNEs and SSIMs of every spatial frame reconstructed by the three methods (Cardiac MR images 1) at 30% sampling rate.

F I G U R E 9 The reconstructed error maps of the 10th spatial frames (Cardiac MR images 1) by the three methods at 30% sampling rate. The colour scale ranges from 0 to 0.1

F I G U R E 10 The reconstructed error maps of the 136th time profiles (Cardiac MR images 1) by the three methods at 30% sampling rate. The colour scale ranges from 0 to 0.1

Considering dynamic MR images have low tubal-rank as shown in Figure 6, we enforce low tubal rank to reconstruct dynamic MR images from highly undersampled k-t space data. This problem can be formulated as below:

$$\min_{X} ||X||_{TN} \quad \text{subject to } R \circ F_{s} X = B$$

(28)

where $X$ denotes a dynamic magnetic resonance image with the size of $n_1 \times n_2 \times n_3$, $F_s$ is the spatial Fourier operator, that is $F_s X$ denotes the Fourier coefficients along the spatial dimension. $R$ is the undersampling tensor mask, which selects 0 or 1 as its elements randomly. $\circ$ denotes Hadamard product. $B$ is the undersampled k-t space measurements with the size of $n_1 \times n_2 \times n_3$. In dynamic MR image reconstruction, the sampling should be different along the $n_3$ dimension of the k-t space.

The two data sets can be obtained from http://hispl.weakly.com/ [43] and http://mri.beckman.illinois.edu/software.html [23] respectively.

The state-of-the-art CS d-MRI methods including L1-TV [44], k-t SLR [20], RPCA-DMRI [45], TuckerDMRI [19] and PS-L1 [23] etc. We have conducted various comparisons of k-t tSVDTV the (t-SVD method combined with sparsity), k-t SLR, RPCA-DMRI, L1-TV, and TuckerDMRI methods in our previous papers [21, 25]. Zero-filled method fills zeros into the undersampled k-space, then gets the reconstructed image $X^r = F_s^{-1} (R \circ F_s X) F_s^{-1} (B)$, where $F_s^{-1}$ is the inverse operator of $F_s$. In most low-rank-based d-MRI methods such as k-t SLR [20], RPCA-DMRI [45] and PS-L1 [23], a 3D dynamic MRI data was unfolded into a 2D matrix, then low matrix-rankness that reflected the inherent spatiotemporal correlation was enforced to reconstruct images. For example, the k-t SLR method first unfolded the 3D dynamic MRI data into a 2D matrix, then low matrix-rankness that reflected the inherent spatiotemporal correlation was enforced to reconstruct images. Simultaneously, it integrated total variation (TV) as sparsity constraint. We call the reconstruction method that only constrains low rankness of the unfolded matrix as unfolding method. Instead of unfolding, tensor decomposition can be used to discover the inherent structural features of dynamic MRI data. In ref. [19], the authors proposed Tucker...
decomposition based dynamic MRI method (TuckerDMRI). TuckerDMRI treated the dynamic images as the sum of a sparse and a low rank component, and then enforced the sparsity of the sparse component and low rankness of the other component by constraining the low rankness of mode-$n$ matrices. We name the method that only constrains the low rankness of mode-n matrices as Tucker method.

We conduct comparisons of three methods exploiting only low-rankness: (1) t-SVD method. (2) Tucker method. (3) Unfold method.

The key parameters of the above methods were chosen empirically by minimising the RLNE of reconstruction results over a range of possible values. In the following experiments of t-SVD method, we set the maximum number of iteration $t_{\text{max}} = 200$, convergence condition $\eta_{\text{tol}} = 10^{-5}$, $\lambda = 1$ and $\rho = 0.003$. The sampling mask and the original d-MRI images used in the following experiments are shown in Figure 7. Experiments were conducted on both complex-valued and real-valued cardiac data sets. The Cardiac MR images 1: Obtained from Bio Imaging and Signal Processing Lab (http://bispl.weebly.com/), this data set contains $N_1 = 25$ temporal frames of size $N_1 = N_2 = 256$ with a $345 \times 270$ mm$^2$ field of view (FOV) and 10 mm slice thickness. The acquisition sequence was steady-state-free precession (SSFP) with a flip angle of 50° and TR = 3.45 ms. The Cardiac MR images 2: A numerical human cardiac MR phantom with quasi-periodic heartbeats provided in http://mri.beckman.illinois.edu/software.html. The phantom was created from real human cardiac MRI data which were collected using retrospective ECG-gating during a single breath-hold and used to generate a time series of images representing a single prototype cardiac cycle [23]. Acquisition parameters: acquisition matrix size = $200 \times 256$, field-of-view (FOV) = $273 \times 50$ mm, effective spatial resolution $1.36 \times 1.36$ mm, slice thickness = $6$ mm, and $TR = 3$ ms.

The RLNEs (%) and SSIMs of the reconstructed images by the four methods with different SRs are shown in Table 2. The RLNEs and SSIMs of every spatial frame under 30% sampling mask for Cardiac MR images 1 reconstruction are shown in Figure 8. The visual comparisons of the reconstructed results by different methods under the 30% sampling mask are shown in Figures 9 and 10. We can see that the Unfold method yields better performances over other methods. The t-SVD method demonstrates better results than Tucker method and comparable results as the unfold method, which verifies that t-SVD framework can be used to reduce the amount of the required sampling data in d-MRI.

The unfold method uses the correlation between spatial frames (frontal slices), which is an effective method for d-MRI, but when all the sampling mask of the frontal slices are the same, as shown in Table 3, the unfold method does work well. Moreover, when the reconstructed data are random tensors with low tubal rank rather than the similar images, the unfold method does not work well either. Our t-SVD method can overcome the above problems. We can see that Tucker method is comparable to t-SVD method when all the sampling mask of the frontal slices are the same. More numerical results related with the t-SVD method can be found in our previous papers [21, 22, 25].

All the above experiments demonstrate that an unknown tensor can be efficiently reconstructed from randomly coefficients by minimising tensor nuclear norm.

Computational complexity: given a 3-order tensor $X_{n_1 \times n_2 \times n_3}$, the computational complexity in Tucker method mainly depends on the nuclear norm minimization of $X_n$. In t-SVD method, it needs the nuclear norm minimization of each block in the block-diagonal matrix. The two methods require computing singular value decompositions (SVD) at each iteration step. The computation burden increases rapidly as matrix sizes and ranks increase. Assuming $n_k = I$ for all $i = 1, 2, 3$, the computational complexity in Tucker method is $O(3I^3)$, while in t-SVD method is $O(3I^3)$. Therefore, the overall complexity is $O(C3I^3)$, $O(C3I^3)$ respectively, where $C$ is the number of iterations.

6 | CONCLUSION

Enforcing low rankness was proved to be effective in sparse sampling. We addressed the problem of applying t-SVD to reconstruct tensors from a small number of coefficients in any given ortho-normal basis.

The contributions are as follows:

1. We proposed a tensor reconstruction model based on t-SVD and gave the theorem and coherence conditions for the tensor reconstruction from limited coefficients in any ortho-normal basis.
2. Specifically, we extended the matrix coherent conditions to tensor coherent conditions and prove that our results hold for any ortho-normal basis meeting the conditions.
3. In the proof, we first validated our results in the case of Fourier-type basis. Then, we generalised the proof to any orthonormal basis using mathematical induction with two modifications.

| Methods      | Cardiac MR images 1 | Cardiac MR images 2 |
|--------------|---------------------|---------------------|
|              | 20%  | 40%  | 60%  | 20%  | 40%  | 60%  |
| Zero-filled  | 41.18/0.6939 | 7.46/0.9757 | 3.62/0.9935 | 36.08/0.6219 | 8.75/0.9177 | 5.38/0.9640 |
| Tucker       | 30.20/0.7946 | 5.79/0.9853 | 3.28/0.9951 | 33.79/0.6401 | 8.14/0.9227 | 4.99/0.9667 |
| Unfold       | 41.18/0.6939 | 7.46/0.9757 | 3.62/0.9935 | 36.08/0.6219 | 8.75/0.9177 | 5.38/0.9640 |
| t-SVD        | 23.17/0.8445 | 7.46/0.9757 | 3.62/0.9935 | 28.50/0.6594 | 8.41/0.9167 | 5.00/0.9647 |

**Table 3** When all the slice of sampling mask are the same: the RLNEs (%)/SSIMs of four methods with different sampling rates.
4. Lastly, we applied t-SVD to reconstruct random tensors and d-MR images from undersampled measurements. In the reconstruction of random tensors, the tensor structure can be maintained successfully. The reconstruction of dynamic MR images demonstrated that t-SVD framework can be used to reduce the amount of the required sampling data.

Nowadays, MR images are commonly multi-channel. But to simplify the application, this work studies the application of our method in the traditional MRI method (single channel). We will consider the application of parallel imaging in future.

ACKNOWLEDGEMENTS
This work was supported by the National Natural Science Foundation of China (No. 61801513), and the basic research project in Information Security Laboratory of National Defense Research and Experiment in 2020 (No. 2020XXAQ02).

REFERENCES
1. Kolda, T.G., Bader, BW: Tensor decompositions and applications. SIAM Rev. 51(3), 455–500 (2009)
2. Long, Z., et al: Low rank tensor completion for multiway visual data. Signal Process. 155, 301–316 (2019)
3. Oseledets, I., Tyrtyshnikov, E.: TT-cross approximation for multidimensional arrays. Lin. Algebra Appl. 432(1), 70–88 (2010)
4. Zhao, Q., et al: Tensor ring decomposition (2016). arXiv:1606.05535
5. Wang, W., Aggarwal, V., Aeron, S.: Efficient low rank tensor ring completion. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5697–5705 (2017)
6. Zhang, Z., Aeron, S.: Exact tensor completion using T-SVD. IEEE Trans. Signal Proc. 65(6), 1511–1526 (2017)
7. Zhang, Z., et al: Novel methods for multilinear data completion and denoising based on tensor SVD. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3842–3849 (2014)
8. Semerci, O., Hao, N., Kilmer, M.E.: Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. IEEE Trans. Image Proc. 23, 1678–1693 (2014)
9. Kilmer, M.E., Borman, K., Hao, N.: Third-order tensors as operators on matrices: a theoretical and computational framework with applications in imaging. SIAM J. Matrix Anal. Appl. 34(1), 148–172 (2013)
10. Das, S., Routray, A., Deb, A.K.: Efficient tensor decomposition approach for estimation of the number of endmembers in a hyperspectral image. J. Appl. Remote Sens. 14(1), 1–17 (2020)
11. Liu, K., et al: Core consistency diagnostic aided by reconstruction error for accurate enumeration of the number of components in PARAFAC models. In: IEEE International Conference on Acoustic Speech and Signal Process, pp. 6635–6639.IEEE (2013)
12. Costa, J.P.C., Haardt, M., Romer, F.: Robust methods based on the HOSVD for estimating the model order in PARAFAC models. In: 5th IEEE Sens. Array and Multichannel Signal Process. Workshop, pp. 510–514.(2008).
13. Zhou, P., et al: Tensor factorization for low-rank tensor completion. IEEE Trans. Image Proc. (Early Access) (2019).
14. Lu, C., et al: Exact low tubal rank tensor recovery from Gaussian measurements. 2018, arXiv:1806.02511v1
15. Lu, C., et al: Tensor robust principal component analysis with A new tensor nuclear norm. IEEE Trans. Pattern Anal. Machine Intell. (Early Access) (2019)
16. Chen, X., et al: Low-rank tensor completion for image and video recovery via capped nuclear norm. IEEE Access. 7, 112142–112153 (2019)
17. Wang, Y., Jiang, Y: Inverse synthetic aperture radar imaging of three-dimensional rotation target based on two-order match Fourier transform. IET Signal Proc. 6(2), 159–169 (2012)
18. Li, Y., et al: Stepped-frequency inverse synthetic aperture radar imaging based on adjacent pulse correlation integration and coherent processing. IET Signal Proc. 5(7), 632–642 (2011)
19. Roohi, S.F., et al: Dynamic MRI reconstruction using low rank plus sparse tensor decomposition. IEEE International Conference on Image Processing (ICIP) (2016)
20. Lingala, S.G., et al: Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR. IEEE Trans. Med Imaging. 30(5), 1042–1054 (2011)
21. Ma, S., Du, H., Mei, W: Dynamic MR image reconstruction from highly undersampled-Space data exploiting low tensor train rank and sparse prior. IEEE Access. 8, 28690–28703 (2020)
22. Ma, S., et al: Dynamic MRI reconstruction exploiting partial separability and t-SVD. In: IEEE 7th Int. Conf. Bioinf. and Comput. Biol., Hangzhou, China, pp. 179–184.(2019)
23. Zhao, B., et al: Image reconstruction from highly undersampled (k, ζ)-Space data with joint partial separability and sparsity constraints, IEEE Trans. Med. Imag. 31(9), 1809–1820 (2012)
24. Han, Y., et al: MR image reconstruction using csupport constraints and group sparsity regularisation. IET Image Proc. 11(3), 155–163 (2017)
25. Ai, J., et al: Dynamic MRI reconstruction using tensor-SVD. 14th IEEE International Conference on Signal Process, Beijing (2018)
26. Gross, D: Recovering low-rank matrices from few coefficients in any basis. IEEE Trans. Inf. Theory. 57(3), 1548–1566 (2011)
27. Braman, K: Third-order tensors as linear operators on a space of matrices. Lin. Algebra Appl. 437(), 1241–1253 (2010)
28. Martin, C.D., Shafer, R., Larse, B: An order-p tensor factorization with applications in imaging. SIAM J. Scientific Comput. 35(1) (2013)
29. Kilmer, M. E., Martin, C.D.: Factorization strategies for third-order tensors. Linear Algebra Appl. 435(3), 641–658 (2011)
30. Singer, A., Cuckerling, M.: Uniqueness of low-rank matrix completion by rigidity theory. SIAM J. Matrix Anal. Appl. 31, 1621–1641 (2010)
31. Candes, E.J., Tao, T: The power of convex relaxation: Near-optimal matrix completion. IEEE Trans. Inf. Theory. 56(5), 2053–2080 (2010)
32. Donoho, D: Compressed sensing. IEEE Trans. Inf. Theor. 52(4), 1289–1306 (2006)
33. Candes, E.J., Romberg, J., Tao, T: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory. 52(2), 489–509 (2006)
34. Candes, E.J., Tao, T: Near-optimal signal recovery from random projections: universal encoding strategies? IEEE Trans. Inf. Theory. 52(12), 5406–5425 (2006)
35. Gross, D, Netse, V: Note on sampling without replacing from a finite collection of matrices (2010). arXiv:1001.2738
36. Cai, J-F., Candes, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. 20(4), 1956–1982 (2010)
37. Yin, M., et al: Subspace clustering via learning an adaptive low-rank graph. IEEE Trans. Image Proc. 27(8), 3716–3728 (2018)
38. Recht, B: A simpler approach to matrix completion. J. Mach. Learn. Res. 12, 3413–3430 (2011)
39. Ahlswede, R., Winter, A: Strong converse for identification via quantum channels. IEEE Trans. Info. Theor. 48(5), 569–579 (2002)
40. Tropp, J.A.: User-friendly tail bounds for sums of random matrices. Found Comput. Math. 12(4), 389–434 (2012)
41. Chen, Y.: Incoherence-optimal matrix completion. IEEE Trans. Inf. Theory. 61(5), 2909–2923 (2015)
42. Wang, Z., et al: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Proc. 13(4), 600–612 (2004)
43. Jung, H., et al: k-t FOCUS: a general compressed sensing framework for high resolution dynamic MRI. Magn. Reson. Med. 61(1), 103–116 (2009)
44. Lustig, M., et al: k-t SPARSE: high frame rate dynamic MRI exploiting spatio-temporal sparsity. Proc. ISMRM. 2420 (2006)
How to cite this article: Ma S, Ai J, Du H, Fang L, Mei W. Recovering low-rank tensor from limited coefficients in any ortho-normal basis using tensor-singular value decomposition. IET Signal Process. 2021;15:162–181. https://doi.org/10.1049/sil2.12017

APPENDICES

Main proof

First, we recall the dual certificate and exact matrix completion problem via convex optimization [46]. For the matrix optimization problem

$$\min \| X \|_s \quad \text{subject to } X_{ij} = M_{ij} \quad (i,j) \in \mathbb{R}_{\Omega}$$ (29)

Candès and Recht [46] showed the duality that if there exists a dual vector $\lambda \in \text{vec}(\mathbb{R}_{\Omega})$ such that $Y = \mathbb{R}_{\Omega} \lambda$ is a sub-gradient of the nuclear norm at the point $X_0$ that is $Y \in \partial \| X \|_s$, then $X_0$ is a solution to Equation (29).

Suppose $X_0 = \sum_{k=1} T N N \sigma_k u_k v_k^T$, the dual certificate $Y$ is a sub-gradient of the convex function $f(X) = \| X \|_s$ at $X_0$ if and only if $Y = \sum_{k=1} T N N \sigma_k u_k v_k^T + W$, where $W$ obeys the two properties:

1) $U^T W = 0$ and $W V = 0$, where $U = [u_1, \ldots, u_r]$, and $V = [v_1, \ldots, v_r]$.
2) $\| W \| \leq 1$.

Similarly, we set $Y$ as the sub-gradient of the convex function $F(X) = \| X \|_{T N N}^T$ at $M_0$, then according to the definition of sub-gradient and Equation (5) we have

$$\| Y \| \geq \| M_0 \|_s + \langle Y, X - M_0 \rangle. \quad (30)$$

So, we can verify that $Y = U * Y^T + W$, where $U * M_0 * Y^T$ is the t-SVD of $M_0$, and $\| W \| \leq 1$ by the duality.

We construct the dual certificate $Y$ satisfies Proposition 1 and show that it is close to the sub-gradient of convex function $\| M \|_{T N N}$ under certain conditions.

Proposition 1 and Lemma 1 directly support the proof of Theorem 2. Also, the following three facts are needed. In Fourier basis case, we present the proof for Proposition 1 and Lemma 1 in Section 4. The relatively simple modifications to cover the general case of Proposition 1 and Lemma 1 are shown in Appendix 7.2. For the proof of the following three facts reader can refer to ref. [6].

**Fact 1** $\| A \|_{T N N} = \rho_3 \text{sup}_{\| B \|_{E U L}}(A, B)$, where $A$ and $B \in \mathbb{R}^{n \times n \times n}$.

Define the t-SVD of $P_{T \perp}(Z)$ to be $P_{T \perp}(Z) = U_\perp * S_\perp * V_\perp^T$, where $Z \in \mathbb{R}^{n \times n \times n}$ such that $P_{T \perp}(Z) = 0$. Then use Fact 1 we have

$$\| P_{T \perp}(Z) \|_{T N N} = n_3 \langle U_\perp * V_\perp^T, P_{T \perp}(Z) \rangle$$

**Fact 2** $\| M \|_{T N N} = n_3 \langle U_\perp * Y^T + U_\perp * V_\perp^T, M \rangle$.

**Fact 3** $\| U * Y^T + U_\perp * V_\perp^T \| = 1$.

Now using the above three facts, Proposition 1 and Lemma 1, given any perturbation $M + Z$ where $P_{T \perp}(Z) = 0$, we have $\| M + Z \|_{T N N} > \| M \|_{T N N}$, which proves $M$ is the unique minimizer of Equation (6). For the proof of $\| M + Z \|_{T N N} > \| M \|_{T N N}$ reader can refer to reference [6]. Here, we give the details process.

Using Fact 1 and Fact 3, and setting $B = U * Y^T + U_\perp * V_\perp^T$ we have:

$$\| M + Z \|_{T N N} \geq n_3 \langle U * Y^T + U_\perp * V_\perp^T, M + Z \rangle \quad (31)$$

Using Fact 2, we have

$$\| M \|_{T N N} = n_3 \langle U * Y^T + U_\perp * V_\perp^T, Z \rangle$$

Using property $Z = P_{T \perp} Z + P_{T \perp} Z$, $\langle U_\perp * V_\perp^T, P_{T \perp}(Z) \rangle = 0$, and $\langle U_\perp * Y^T, P_{T \perp}(Z) \rangle = 0$, we have

$$\| M \|_{T N N} = n_3 \langle U_\perp * Y^T, P_{T \perp}(Z) \rangle + n_3 \langle U_\perp * V_\perp^T, P_{T \perp}(Z) \rangle \quad (32)$$

Introducing the tensor dual certificate $Y$ supported in $\Omega$ such that $P_{T \perp}(Y) = Y$. Since $P_{T \perp}(Z) = 0$, it is easy to show $(Z, Y) = 0$. Then, we have

$$= \| M \|_{T N N} + n_3 \langle U_\perp * Y^T, P_{T \perp}(Z) \rangle$$

Using Definition 8, Fact 2 and Remark 1, we have
In this appendix, we show that the construction of the conditions (Equation 12 and 13). As shown in Section 4, Golfing scheme is used in the proof of Lemma 2. Before we give the proof for Proposition 1(1) and Lemma 1 in Section 4 are validated in both type of basis case. So, next we only need to present the proof of Lemma 3 and Lemma 4.

**Proof of Lemma 3:** Observe that

\[
\mu((P_T - P_T R_{\omega} P_T)Z) = \max_a (W_a, (P_T - P_T R_{\omega} P_T)Z)^2
\]

\[
\text{Fix } b \in [1, n^2 n_3], \quad Z \in T, \quad P_T Z = Z, \quad \text{and } R_{\omega} Z = \sum_a n \frac{1}{p} \delta_a(Z, W_a) W_a, \quad \text{we get}
\]

\[
(W_b, (P_T - P_T R_{\omega} P_T)Z) = (W_b, Z) - (W_b, P_T R_{\omega} P_T Z)
\]

\[
= (W_b, Z) - \sum_a \frac{1}{p} \delta_a(Z, P_T W_a) (W_b, P_T W_a)
\]

Define

\[
\mathcal{L}_a(Z) = \frac{1}{n^2 n_3} (W_b, Z) - \frac{1}{p} \delta_a(Z, P_T W_a) (W_b, P_T W_a)
\]

We have

\[
E[L_a(Z)] = \frac{1}{n^2 n_3} E[(W_b, Z)] - \frac{1}{n^2 n_3} E \left[ \sum_a \frac{1}{p} \delta_a(Z, P_T W_a) (W_b, P_T W_a) \right]
\]

\[
= \frac{1}{n^2 n_3} E[(W_b, Z)] - \frac{1}{n^2 n_3} E[(W_b, P_T R_{\omega} Z)]
\]

where \(z \geq \|Z\|_2\) is an upper bound on the 2-norm of \(Z\). By assumption Equation (13), the estimate Equation (36) holds during the process of Golfing scheme.

**Lemma 4** If \(p\) satisfies the condition in Theorem 2, and \(Z \in T\) sized \(n \times n \times n_3\) with an upper bound \(z\) on its 2-norm. Then for some constants \(c_5\), we have

\[
\|P_T R_{\omega} Z\| \leq \frac{1}{4} \sqrt{z}
\]
In the Bernoulli model, $\mathbb{E}[R_{ij}] = I$. So it is obvious to know that $\mathbb{E}L_a(Z) = 0$. Using assumption (Equation 12) we consider that

$$\|\mathcal{L}_a\| = \|L_a\| \leq \max \left\{ \frac{1}{n^2n_3} \mu(Z), \frac{1}{p} \|P_TW_a\|^2 \right\} \leq \frac{2\nu r}{pn} \mu(Z)$$

Using assumption (Equation 12), we also have

$$\left\| \mathbb{E} \left[ \sum_a \mathcal{L}_a^2 \right] \right\| = \left\| \mathbb{E} \left[ \sum_a L_a^2 \right] \right\| = \left\| \sum_a \left( \frac{1}{n^2n_3} (W_b, Z)^2 + \frac{1}{p} \|P_TW_a\|^2 \langle W_b, P_TW_a \rangle \right) \right\| \leq \frac{1}{n^2n_3} \mu(Z) + \frac{1}{p} \|P_TW_a\|^2 \mu(Z) \leq \frac{1}{n^2n_3} \mu(Z)$$

Then use Theorem 3, we have

$$\mathbb{P} \left[ \langle W_b, (P_T - P_T R_a P_T) Z \rangle > \frac{1}{2} (W_b, Z) \right] \leq (2n_3) \exp \left( \frac{-1}{8} \mu(Z) \right) \leq (2n_3)^{-c'_4}$$

for some $c'_4$ large enough.

The advertised estimate follows by taking squares and applying the union bound over the $n^2n_3$ elements of the basis.

Proof of Lemma 4: Similar to the proof of Lemma 2, we need to consider the quantity of Equation (36)

$$\mathcal{H}_a(Z) = \frac{1}{p} \delta_b (W_a, Z) P_T = W_a$$

Assume that $Z \in T$ with $\|Z\|_2 = 1$, then using the estimate Equation (36)

$$\max_a |\langle W_a, Z \rangle|^2 \leq \frac{\nu}{n^4n_3}$$

holds. First note that

$$\mathbb{E} \left[ \sum_a \mathcal{H}_a^2(Z) \right] = \mathbb{E} \left[ \sum_a \mathcal{H}_a^2(Z) \right] \leq \frac{1}{p} \sum_a \langle W_a, Z \rangle^2 \langle P_T W_a \rangle^2 \leq \frac{1}{p} \|P_TW_a\|^2 \langle \psi, W_a^2 \psi \rangle$$

where the maximum is over all normalised columns $\psi \in (\text{range}M)^\perp$ sized $n \times 1 \times n_3$. Let $\psi_0$ be a column achieving the maximum. Define two vectors $x, y \in \mathbb{R}^{n^2n_3}$ by setting their components to

$$x_a := \langle W_a, Z \rangle, \quad y_a := \frac{1}{n^2n_3} \langle \psi_0, W_a^2 \psi_0 \rangle$$

respectively. The assumption that $\|Z\|_2 = 1$ implies that $\|x\|_1 = \|y\|_1 = 1$. In addition, the same is true for the other vector: $\|y\|_1 = 1$, regardless of the basis chosen. So we can compute

$$\left( \sum_a W^T_i W_i \right)_{i,j} = \sum_{a,k} \langle W_a \rangle_{k,j} : \langle W_a \rangle_{k,j} = n^2n_3 e_{i,j}$$

where $e_{i,j} \in \mathbb{R}^{1 \times 1 \times n_3}$ is a tensor tube.

Thus $\|y\|_1 = \sum_a y_a = \frac{1}{n^2n_3} \langle \psi_0, n^2n_3 \psi_0 \rangle = 1$

$$\|x\|_1 \leq \min \{ \|x\|_1, \|y\|_1 \} \leq \frac{\nu}{n^4n_3}$$

Then we get

$$\mathbb{E} \left[ \sum_a \mathcal{H}_a^2(Z) \right] \leq \mathbb{E} \left[ \sum_a \mathcal{H}_a^2(Z) \right] \leq \frac{\mathbb{E}}{pn^2} \nu \leq \frac{\nu}{pn}$$

Similarly,

$$\mathcal{H}_a(Z) = \mathbb{E} |\mathcal{H}_a(Z)| \leq \frac{1}{p} \left\| \langle W_a, Z \rangle^2 \right\| \leq \frac{1}{p} \sqrt{n^2n_3 \|x\|_1} \leq \frac{1}{pn} \sqrt{\nu} \leq \frac{\sqrt{\nu}}{pn}$$
Then use Theorem 3, we have

\[
\Pr \left[ \left\| \mathcal{P}_T \mathcal{R}_\Omega Z \right\|_F > \frac{1}{4 \sqrt{r}} \right] 
\leq (2m_3)^{-c_3} \cdot \exp \left( \frac{-1}{32 r} \frac{z^2}{\nu} \right) \leq (2m_3)^{-c_3}
\]

for some \( c_3 \) large enough, which finish the proof of the general case.

The Golging scheme is first introduced in ref. [26]. It was used to construct the matrix dual certificate iteratively in ref. [26]. We adopt the Golging scheme in our proof to construct a dual certificate tensor \( \mathcal{Y} \) and then prove that it satisfies both conditions of Proposition 1(2). The explanation of Golging scheme used is provided below.

Let \( \Omega \) be a union of smaller sets \( \Omega_t \) such that \( \Omega = \bigcup_{t=1}^{t_0} \Omega_t \), where for each \( t \) and tensor index \( (i, j, k) \in \Omega_t \), the probability \( \mathbb{P}(i, j, k) \in \Omega_t \) is independent of all others. \( t_0 \) is the total number of iteration. We assume order-3 tensors \( \mathcal{G}_0, \mathcal{G}_1, \ldots, \mathcal{G}_{t_0} \) can be obtained by the following iteration.

\[
\begin{align*}
\mathcal{G}_0 &= 0 \\
\mathcal{G}_1 &= \mathcal{R}_\Omega (\mathcal{Z}) \\
\mathcal{G}_2 &= \mathcal{G}_1 + \mathcal{R}_\Omega (\mathcal{Z} - \mathcal{P}_T (\mathcal{G}_1)) \\
dots & \\
\mathcal{G}_{t_0} &= \mathcal{G}_{t_0-1} + \mathcal{R}_\Omega (\mathcal{Z} - \mathcal{P}_T (\mathcal{G}_{t_0-1}))
\end{align*}
\]

where tensor \( \mathcal{Z} \) meeting the following “distance” criteria

\[
\| \mathcal{Z} - \mathcal{P}_T (\mathcal{G}_t) \|_F \leq \frac{1}{2} \| \mathcal{Z} - \mathcal{P}_T (\mathcal{G}_{t-1}) \|_F,
\]

for each \( t \). We set the dual certificate tensor \( \mathcal{Y} = \mathcal{G}_{t_0} \). Then, applying the above inequality recursively, we have

\[
\| \mathcal{P}_T (\mathcal{Y}) - \mathcal{Z} \|_F \leq \left( \frac{1}{2} \right)^{t_0} \| \mathcal{Z} \|_F.
\]

As \( t_0 \to +\infty \), \( \mathcal{P}_T (\mathcal{Y}) \) converges exponentially fast to the original tensor \( \mathcal{Z} \). For a more intuitive presentation, we give the figure of \( \mathcal{G}_t \) changing with \( t \). As shown in Figure 11, \( \mathcal{G}_t \) moves forward slightly and converges exponentially fast to the origin \( \mathcal{G}_0 = \mathcal{R}_\Omega (\mathcal{Z}) \) with small “moving distance” \( \| \mathcal{Z} - \mathcal{P}_T (\mathcal{G}_t) \|_F \). Because it’s like playing golf, D. Gross called it the Golging scheme.