Tensor Matched Subspace Detection
Cuiping Li, Xiao-Yang Liu, and Yue Sun

Abstract
The problem of testing whether an incomplete tensor lies in a given tensor subspace is significant when it is impossible to observe all entries of a tensor. We focus on this problem under tubal-sampling and elementwise-sampling. Different from the matrix case, our problem is much more challenging due to the curse of dimensionality and different definitions for tensor operators. In this paper, the problem of matched subspace detections are discussed based on tensor product (t-product) and tensor-tensor product with invertible linear transformation (L-product). Based on t-product and L-product, a tensor subspace can be defined, and the energy of a tensor outside the given subspace (also called residual energy in statistics) is bounded with high probability based on samples. For a tensor in \( \mathbb{R}^{n_1 \times 1 \times n_3} \), the reliable detection is possible when the number of its elements we obtained is slightly greater than \( r \times n_3 \) both with t-product and L-product, where \( r \) is the dimension of the given tensor subspace.

Index Terms
Tensor subspace detection, elementwise-sampling, tubal-sampling, residual energy, linear transforms, L-product.

I. INTRODUCTION
Matched subspace detection is widely used in many applications such as hyperspectral target detection [1], image representation [2], MIMO system [3, 4], shape detection [5], etc. Conventional approaches of matched subspace detection are based on the algebra of vectors and matrices [6, 7], which can not work well when the data is represented as a tensor. Therefore it is urgent to proposed a more efficient approaches of matched subspace detection.

In [1], a tensor matched subspace detector based on the \( n \)-mode products is proposed. In this framework, although the spatial-spectral information of signals has been taken into account, the computation of the orthogonal matrix onto the given subspace is complex.

We are motivated to propose an approach for tensor matched subspace detection for the following reason:
• Tensors as mathematical tools are widely used in signal processing and data analysis. And tensor represented signals contain more information than vector represented signals.

• The t-product has been introduced as a useful multiplication between third-order tensors, and under this framework, the projection is well defined.

In this paper, we proposed a method for tensor matched subspace detection based on t-product. Compare with tensor matched subspace detector in [1], our approach is more practical and concise. Furthermore, our approach has more accurate residual energy bound than that in [1].

The remainder of this paper is organized as follows. In Section II, the notations and definitions based on t-product are introduced. In Section III, we present the problem of tensor matched subspace detection and main theorems of this paper. Section IV provides the hypothesis testing for tensor matched subspace detection, both with and without noise. In Section V, tensor matched subspace detection with invertible linear transformation is discussed.

II. NOTATIONS AND PRELIMINARIES

Vectors are denoted by lowercase letters, e.g., \( a \); tubal-scalars are denoted by boldface lowercase letters, e.g., \( \mathbf{a} \); matrices are denoted by capital letters, e.g., \( A \); and third-order tensors are denoted by calligraphic letters, e.g., \( A \). The transpose of a vector or a matrix is denoted with a superscript \( T \), i.e., \( a^T \), \( A^T \). The \( i \)th element of a vector \( a \) is \( a_i \), the \( (i,j) \)th element of a matrix \( A \) is \( A_{ij} \) or \( A(i,j) \), and the \( (i,j,k) \)th element of a third-order tensor \( A \) is \( A_{ijk} \) or \( A(i,j,k) \). We use \( [n] \) to denote the index set \( \{1,2,\ldots,n\} \), and \( [n_1] \times [n_2] \) to denote the set \( \{(1,1),(1,2),\ldots,(1,n_2),(2,1),\ldots,(n_1,n_2)\} \).

**Tubes and slices of a tensor**[8]: For a third-order tensor \( A \in \mathbb{R}^{n_1 \times n_2 \times n_3} \), a tube of \( A \) is defined by fixing all indices but one, while a slice of \( A \) defined by fixing all but two indices. We use \( A(:,j,:) \), \( A(i,:) \), \( A(i,j,:) \) to denote mode-1, mode-2, mode-3 tubes of \( A \), and \( A(:,:,:), A(:,j,:), A(i,:,:), A(i,:,:) \) to denote the frontal, lateral, and horizontal slices of \( A \). \( A(i,:,:) \) and \( A(:,j,:) \) are also called tensor row and tensor column. For easy representation, we use \( A(k,:) \) to denote \( A(:,:,k) \).

**Definition 1** (t-product [8, 9]). The tensor-product \( C = A \ast B \) of \( A \in \mathbb{R}^{n_1 \times n_2 \times k} \) and \( B \in \mathbb{R}^{n_2 \times n_3 \times k} \) is a tensor of size \( n_1 \times n_3 \times k \) with \( C(i,j,:) = \sum_{s=0}^{n_2} A(i,s,:) \ast B(s,j,:) \), for \( i \in [n_1] \) and \( j \in [n_3] \), where \( A(i,s,:) \ast B(s,j,:) \) denotes the circular convolution between two tubes.

For the convenience of the next discussion, we give a introduction of bcirc(·) and unfold(·) operations [9]. The block circular matrix representation of tensor \( A \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) is denoted by bcirc(\( A \), as
follows
$$\text{bcirc}(\mathcal{A}) = \begin{bmatrix}
\mathcal{A}^{(1)} & \mathcal{A}^{(n_3)} & \cdots & \mathcal{A}^{(2)} \\
\mathcal{A}^{(2)} & \mathcal{A}^{(1)} & \cdots & \mathcal{A}^{(3)} \\
\vdots & \vdots & \ddots & \vdots \\
\mathcal{A}^{(n_3)} & \mathcal{A}^{(n_3-1)} & \cdots & \mathcal{A}^{(1)}
\end{bmatrix}.$$ 

The operation of unfold(·) on an $n_1 \times n_2 \times n_3$ tensor $\mathcal{A}$ is defined as follows
$$\text{unfold}(\mathcal{A}) = \begin{bmatrix}
\mathcal{A}^{(1)} \\
\mathcal{A}^{(2)} \\
\vdots \\
\mathcal{A}^{(n_3)}
\end{bmatrix}.$$ 

And we define fold(·) operation as the inverse operation of unfold(·). Based on these operations, the t-product between $\mathcal{A}$ and $\mathcal{B}$ can be rewritten as follows:
$$\mathcal{A} \ast \mathcal{B} = \text{fold}(\text{bcirc}(\mathcal{A}) \cdot \text{unfold}(\mathcal{B})).$$

**Definition 2** (Tensor transpose [9]). Let $\mathcal{A}$ be an $n_1 \times n_2 \times n_3$ tensor, then $\mathcal{A}^\dagger$ is the $n_2 \times n_1 \times n_3$ tensor obtained by transposing each frontal slice and then reversing the order of transposed frontal slices 2 through $n$, i.e., $\mathcal{A}^\dagger(:,:,1) = \mathcal{A}(::,:,1)^\dagger$ and $\mathcal{A}^\dagger(:,:,k) = \mathcal{A}(:::,n_3 + 2 - k)^\dagger$, $2 \leq k \leq n_3$.

**Definition 3** (Identity tensor [8, 9]). The identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times k}$ is a tensor whose first frontal slice $\mathcal{I}(::,1)$ is an $n \times n$ identity matrix and all other frontal slices $\mathcal{I}(::,i)$, $(i = 2, \ldots, k)$ are zero matrices.

**Definition 4** (Inverse [8, 9]). The inverse of a tensor $\mathcal{A} \in \mathbb{R}^{n \times n \times k}$ is written as $\mathcal{A}^{-1} \in \mathbb{R}^{n \times n \times k}$ and satisfies $\mathcal{A}^{-1} \ast \mathcal{A} = \mathcal{A} \ast \mathcal{A}^{-1} = \mathcal{I}$.

**Definition 5** (Orthogonal tensor [8, 9]). A tensor $\mathcal{A} \in \mathbb{R}^{n \times n \times k}$ is orthogonal if it satisfies $\mathcal{A}^\dagger \ast \mathcal{A} = \mathcal{I}$.

**Definition 6** (Tensor linear combination [9]). Given $k$ tubal scalars $\mathbf{c}_j$, $j \in [k]$, in $\mathbb{K}_n$, a t-linear combination of $\mathcal{X}_j \in \mathbb{R}^{m \times 1 \times n}$, $j \in [k]$, is defined as
$$\mathcal{X}_1 \ast \mathbf{c}_1 + \mathcal{X}_2 \ast \mathbf{c}_2 + \cdots + \mathcal{X}_k \ast \mathbf{c}_k \equiv \mathcal{X} \ast \mathcal{C}, \text{ where } \mathcal{X} \triangleq [\mathcal{X}_1, \ldots, \mathcal{X}_k], \mathcal{C} \triangleq \begin{bmatrix}
\mathbf{c}_1 \\
\vdots \\
\mathbf{c}_k
\end{bmatrix}.$$ 

**Definition 7** (f-diagonal tensor [8, 9]). A tensor is called f-diagonal tensor if each frontal slice of the tensor is a diagonal matrix.
Definition 8 (t-SVD [8, 9]). The t-SVD of $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is given by $A = U \Sigma V^\dagger$, where $U$ and $V$ are orthogonal tensors of size $n_1 \times n_1 \times n_3$ and $n_2 \times n_2 \times n_3$ respectively, and $\Sigma$ is a f-diagonal tensor of size $n_1 \times n_2 \times n_3$. The entries of $\Sigma$ are called the singular values of $A$.

Definition 9 (Tensor tubal-rank [8, 9]). The tensor tubal-rank of a third-order tensor $A$ is the number of none-zero tubal-scalars of $\Sigma$ in the t-SVD.

Definition 10 (Tensor-column subspace). If $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with tubal-rank of $r$, the $r$-dimensional tensor-column subspace $S$ spanned by columns of $A$ is defined as

$$S = \{X | X = A_1 \ast c_1 + A_2 \ast c_2 + \cdots + A_{n_2} \ast c_{n_2}\},$$

where $c_j, j \in [n_2]$, is an arbitrary tubal scalar of length $n_3$.

Let $S$ be the subspace spanned by columns of third-order tensor $A$. When $A^\dagger \ast A$ is invertible, $A \ast (A^\dagger \ast A)^{-1} \ast A^\dagger$ is an orthogonal projection onto $S$ [9].

Frequency domain representation [8]: $\hat{A}$ denotes a third-order tensor obtained by taking the Fourier Transform of all the tubes along the third dimension of an $n_1 \times n_2 \times n_3$ tensor $A$, i.e., for $i \in [n_1]$ and $j \in [n_2]$, $\hat{A}(i, j, :) = \text{fft}(A(i, j, :))$. And we use $\mathcal{A}$ to denote the block-diagonal matrix of the tensor $\hat{A}$ in the Fourier domain, i.e.,

$$\mathcal{A} = \begin{bmatrix} \hat{A}^{(1)} & & \\ & \hat{A}^{(2)} & \\ & & \ddots \\ & & & \hat{A}^{(n_3)} \end{bmatrix}.$$ 

Tensor norm [8–11]: For a tubal scalar $a \in \mathbb{R}^{1 \times 1 \times n_3}$, we define an $\ell_2$ norm on it as $\|a\|_2 = \sum_{k=1}^{n_3} a_k^2$, where $a_k$ is the $k$th entry of $a$. For an $n_1 \times n_2 \times n_3$ tensor $A$, its infinity norm is defined as $\|A\|_\infty = \max_{i \in [n_1]} \max_{j \in [n_2]} |A_{ijk}|$. Its Frobenius norm is defined as $\|A\|_F = \sqrt{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} A_{ijk}^2}$, and $\|A\|_F = \frac{1}{\sqrt{n_3}} \|\hat{A}\|_F$ [11]. Its tensor spectral norm $\|A\|_2$ is defined as the largest singular value of $A$ [11], and $\|A\|_2 = \|bcirc(A)\|_2 = \|A\|_2$, where $\|bcirc(A)\|_2$ and $\|A\|_2$ denote the matrix spectral norm. For a tensor column $X \in \mathbb{R}^{n_1 \times 1 \times n_3}$, we define an $\ell_\infty^*$ norm as $\|X\|_\infty^* = \max_i \|X(i, 1, :)\|_2$.

III. PROBLEM STATEMENT AND MAIN THEOREMS

In this section, we first describe the problem of tensor matched subspace detection, and then we give two main theorems which are significant for us to decided whether a signal belongs to a given subspace with high probability.
A. Problem Statement

Let $V \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ denote a signal with its entries are sampled with replacement, and the given subspace $S$ is spanned by the columns of $U \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, hence $P_S = U \ast (U^\dagger \ast U)^{-1} \ast U^\dagger$ is the orthogonal projection onto $S$. Then we want to find out how many samples are required to decided whether a signal belongs to a given subspace with high probability. In order to solve this problem, we must quantify how much information each sample can provide. We define the coherence of a subspace $S$ according to the definition of the tensor incoherence condition in [8] to be the quantity

$$\mu(S) \triangleq \frac{n_1}{r} \max_j \|P_S \ast \hat{e}_j\|_F^2,$$

where $r$ is the dimension of $S$ and $\hat{e}_j$ is a tensor column basis of size $n_1 \times 1 \times n_3$ with $\hat{e}_{j1} \equiv 1$ and the rest entries equaling zero [8].

Let $V = X + Y$, where $X \in S$ and $Y \in S^\perp$, and the dimension of $S$ is $n_2$. And we consider two types of sampling: tubal-sampling and elementwise-sampling.

**Tubal-sampling:** Let $V$ be a vector of $n_1$ with elements being tubes. Let $\Omega_1$ be the index set of samples and $\Omega_1 \subset [n_1]$. If $i \in \Omega_1$, $V(i, 1, :)$ is a sample.

**Elementwise-sampling:** Let $\Omega_2$ be the index set of samples and $\Omega_2 \subset [n_1] \times [n_3]$, i.e., if $(i, j) \in \Omega_2$, $V(i, 1, j)$ is a sample.

Let $\Omega \in \{\Omega_1, \Omega_2\}$, and $V_\Omega$ denote the sampling signal. Let $P_{S_\Omega} = U_\Omega \ast (U_\Omega^\dagger \ast U_\Omega)^{-1} \ast U_\Omega^\dagger$ where $U_\Omega$ is relevant to $U$ and $\Omega$ such that $\|V_\Omega - P_{S_\Omega} \ast V_\Omega\|_2^2 = \|V_\Omega - P_{S_{\Omega_1}} \ast V_{\Omega_1}\|_2^2$. Let $m = |\Omega|$, where $|\Omega|$ denotes the cardinality of $\Omega$, thus $m \in \{m_1, m_2\}$ where $m_1 = |\Omega_1|$ and $m_2 = |\Omega_2|$. And we hope to determine the value of $m$ so that we can decide whether $V$ belongs $S$ based on $V_\Omega$ with high probability.

B. Main Theorem under Tubal-sampling

When the sampling method is tubal-sampling, $|\Omega| = |\Omega_1|$, and $m = m_1$. Then $V_{\Omega_1} \in \mathbb{R}^{m_1 \times 1 \times n_3}$ denotes the sampling signal with its entries satisfied

$$V_{\Omega_1}(i, 1, :) = V(i, 1, :), \ i \in [\Omega_1].$$

Let $P_{S_{\Omega_1}} = U_{\Omega_1} \ast (U_{\Omega_1}^\dagger \ast U_{\Omega_1})^{-1} \ast U_{\Omega_1}^\dagger$ as the projection where $U_{\Omega_1}$ satisfies

$$U_{\Omega_1}(i, :, :) = U(i, :, :), \ i \in [\Omega_1].$$

Then we have the following theorem.
Theorem 1. Let $\delta > 0$ and $m_1 \geq \frac{n_1}{2} n_2 \mu(S) \ln\left(\frac{2n_3 n_5}{\delta}\right)$. Then with probability at least $1 - 4\delta$,

$$\frac{m_1 (1 - \alpha)}{n_1} - n_2 \mu(S) \frac{\beta}{(1 - \gamma)} \|V - \mathcal{P}_S \ast V\|_F^2 \leq \|\mathcal{V}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{V}_{\Omega_1}\|_F^2 \leq (1 + \alpha) \frac{m_1}{n_1} \|V - \mathcal{P}_S \ast V\|_F^2 \tag{4}$$

holds, where $\alpha = \sqrt{2(n_1 \|Y\|_F^2 - \|Y\|_F^2) \ln\left(\frac{1}{\delta}\right)} + \frac{2(n_1 \|Y\|_F^2 - \|Y\|_F^2)}{3m_1 \|Y\|_F^2} \ln\left(\frac{1}{\delta}\right)$, $\beta = \left(1 + 2 \sqrt{\ln\left(\frac{1}{\delta}\right)}\right)^2$, and $\gamma = \sqrt{8n_2 \mu(S)} \frac{3m_1}{n_1} \ln\left(\frac{2n_3 n_5}{\delta}\right)$.

Proof: In order to prove Theorem 1, we first give three Lemmas whose proofs are provided in Appendix.

Lemma 1. With the same notations as Theorem 1,

$$(1 - \alpha) \frac{m_1}{n_1} \|Y\|_F^2 \leq \|\mathcal{Y}_{\Omega_1}\|_F^2 \leq (1 + \alpha) \frac{m_1}{n_1} \|Y\|_F^2 \tag{5}$$

holds with probability at least $1 - 2\delta$.

Lemma 2. With the same notations as Theorem 1,

$$\left\|U_{\Omega_1}^\dagger \ast \mathcal{Y}_{\Omega_1}\right\|_F^2 \leq \beta \frac{m_1 n_2 \mu(S)}{n_1^2} \|Y\|_F^2 \tag{6}$$

holds with probability at least $1 - \delta$.

Lemma 3. With the same notations as Theorem 1,

$$\left\|\left(U_{\Omega_1}^\dagger \ast U_{\Omega_1}\right)^{-1}\right\|_2 \leq \frac{n_1}{(1 - \gamma) m_1} \tag{7}$$

holds with probability at least $1 - \delta$, and $\gamma < 1$.

Recall $\mathcal{V} = \mathcal{X} + \mathcal{Y}$ where $\mathcal{X} \in \mathcal{S}$ and $\mathcal{Y} \in \mathcal{S}^\perp$. Hence $\|\mathcal{V}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{V}_{\Omega_1}\|_F^2 = \|\mathcal{V}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{Y}_{\Omega_1}\|_F^2$.

We split $\|\mathcal{V}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{Y}_{\Omega_1}\|_F^2$ into three terms and bound each with high probability.

$$\|\mathcal{V}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{Y}_{\Omega_1}\|_F^2 = \left(\mathcal{V}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{Y}_{\Omega_1}\right)^\dagger \ast \left(\mathcal{V}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{Y}_{\Omega_1}\right) \tag{1}$$

$$= \left(U_{\Omega_1}^\dagger \ast \mathcal{Y}_{\Omega_1} - \mathcal{Y}_{\Omega_1}^\dagger \ast \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{Y}_{\Omega_1}\right) \tag{1}$$

$$= \|\mathcal{Y}_{\Omega_1}\|_F^2 - \left(U_{\Omega_1}^\dagger \ast \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{Y}_{\Omega_1}\right) \tag{1}$$

$$= \|\mathcal{Y}_{\Omega_1}\|_F^2 - \left(U_{\Omega_1}^\dagger \ast \mathcal{Y}_{\Omega_1} - \left(U_{\Omega_1}^\dagger \ast U_{\Omega_1}\right)^{-1} \ast U_{\Omega_1}^\dagger \ast \mathcal{Y}_{\Omega_1}\right) \tag{1}. \tag{8}$$
Let $W_{\Omega_1} \star W_{\Omega_1} = \left( U_{\Omega_1}^T \star U_{\Omega_1} \right)^{-1}$. Then we have

$$\left( \mathcal{Y}_{\Omega_1}^T \star U_{\Omega_1} \star \left( U_{\Omega_1}^T \star U_{\Omega_1} \right)^{-1} \star U_{\Omega_1} \star \mathcal{Y}_{\Omega_1} \right)^{(1)} = \left\| W_{\Omega_2} \star U_{\Omega_1} \star \mathcal{Y}_{\Omega_1} \right\|_F^2 \leq \left\| \text{bcirc}(W_{\Omega_1}) \right\|_2^2 \left\| U_{\Omega_1} \star \mathcal{Y}_{\Omega_1} \right\|_F^2 = \left\| \text{bcirc} \left( W_{\Omega_1} \star W_{\Omega_1} \right) \right\|_2 \left\| U_{\Omega_1} \star \mathcal{Y}_{\Omega_1} \right\|_F^2 = \left\| \text{bcirc} \left( \left( U_{\Omega_1}^T \star U_{\Omega_1} \right)^{-1} \right) \right\|_2 \left\| U_{\Omega_1} \star \mathcal{Y}_{\Omega_1} \right\|_F^2 = \left\| \left( U_{\Omega_1}^T \star U_{\Omega_1} \right)^{-1} \right\|_2 \left\| U_{\Omega_1} \star \mathcal{Y}_{\Omega_1} \right\|_F^2.$$

Thus

$$\left\| \mathcal{Y}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \star \mathcal{Y}_{\Omega_1} \right\|_F^2 \geq \left\| \mathcal{Y}_{\Omega_1} \right\|_F^2 - \left\| \left( U_{\Omega_1}^T \star U_{\Omega_1} \right)^{-1} \right\|_2 \left\| U_{\Omega_1} \star \mathcal{Y}_{\Omega_1} \right\|_F^2,$$

and combining Lemma 1, Lemma 2, and Lemma 3, we have

$$\frac{m_1(1 - \alpha) - n_2 \mu(S) T_{\beta}}{n_1} \left\| \mathcal{Y} \right\|_F^2 \leq \left\| \mathcal{Y}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \star \mathcal{Y}_{\Omega_1} \right\|_F^2 \leq (1 + \alpha) \frac{m_1}{n_1} \left\| \mathcal{Y} \right\|_F^2$$

with probability at least $1 - 4\delta$.

### C. Main Theorem under Elementwise-sampling

When the sampling method is elementwise-sampling, $|\Omega| = |\Omega_2|$, and $m = m_2$. Then $\mathcal{Y}_{\Omega_2} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ denotes the sampling signal with its entries satisfied

$$\mathcal{V}_{\Omega_2}(i, 1, j) = \begin{cases} \mathcal{V}(i, 1, j), & (i, j) \in \Omega_2; \\ 0, & \text{otherwise}. \end{cases} (10)$$

Let $v = \text{unfold}(\mathcal{V})$ and $U = \text{bcirc}(\mathcal{U})$, then define $\mathcal{P}_{S_{\Omega_2}} = U_{\Omega_2} (U_{\Omega_2}^T U_{\Omega_2})^{-1} U_{\Omega_2}^T$ as the projection where $U_{\Omega_2}$ satisfies

$$U_{\Omega_2}((j - 1)n_1 + i, :) = \begin{cases} U((j - 1)n_1 + i, :) , & (i, j) \in \Omega_2; \\ 0, & \text{otherwise}. \end{cases} (11)$$

And we have the following theorem.

**Theorem 2.** Let $\delta > 0$ and $m_2 \geq \frac{8}{\gamma} n_2 n_3 \mu(S) \ln(\frac{2n_1 n_3}{\gamma})$. Then with probability at least $1 - 4\delta$

$$\frac{m_2(1 - \alpha) - n_2 \mu(S) T_{\beta}}{n_1 n_3} \left\| \mathcal{V} - \mathcal{P}_S \star \mathcal{V} \right\|_F^2 \leq \left\| v_{\Omega_2} - \mathcal{P}_{S_{\Omega_2}} v_{\Omega_2} \right\|_2^2 \leq (1 + \alpha) \frac{m_2}{n_1 n_3} \left\| \mathcal{V} - \mathcal{P}_S \star \mathcal{V} \right\|_F^2$$

holds, where $\alpha = \sqrt{\frac{2(n_1, n_2) \left\| \mathcal{V} \right\|_F^2}{m_2 \left\| \mathcal{V} \right\|_F^2}} \ln(\frac{1}{\delta}) + \frac{2(n_1, n_2) \left\| \mathcal{V} \right\|_F^2 - \left\| \mathcal{V} \right\|_F^2}{3m_2 \left\| \mathcal{V} \right\|_F^2} \ln(\frac{1}{\delta})$, $\beta = \left( 1 + 2\sqrt{\ln(\frac{1}{\delta})} \right)^2$, and

$$\gamma = \sqrt{\frac{8n_2 n_3 \mu(S)}{3m_2} \ln(\frac{2n_1 n_3}{\delta})}.$$
\textbf{Proof:} In order to prove Theorem 2, we must give another three Lemmas whose proofs are provided in Appendix.

\textbf{Lemma 4.} With the same notations as Theorem 2,
\begin{equation}
(1 - \alpha)\frac{m_2}{n_1 n_3} \|Y\|_F^2 \leq \|Y_{\Omega_2}\|_F^2 \leq (1 + \alpha)\frac{m_2}{n_1 n_3} \|Y\|_F^2
\end{equation}
holds with probability at least \(1 - 2\delta\).

\textbf{Lemma 5.} With the same notations as Theorem 2,
\begin{equation}
\|U_{\Omega_2}^T y_{\Omega_2}\|_2^2 \leq \beta \frac{m_2 n_2}{n_1 n_3} \mu(S) \|Y\|_F^2
\end{equation}
holds with probability at least \(1 - \delta\).

\textbf{Lemma 6.} With the same notations as Theorem 2,
\begin{equation}
\left\|\left(U_{\Omega_2}^T U_{\Omega_2}\right)^{-1}\right\|_2 \leq \frac{n_1 n_3}{(1 - \gamma)m_2}
\end{equation}
holds with probability at least \(1 - \delta\), and \(\gamma < 1\).

Recall \(V = \mathcal{X} + \mathcal{Y}\) again where \(\mathcal{X} \in \mathcal{S}\) and \(\mathcal{Y} \in \mathcal{S}^\perp\), hence \(\|v_{\Omega_2} - P_{S_{\Omega_2}} v_{\Omega_2}\|_2^2 = \|y_{\Omega_2} - P_{S_{\Omega_2}} y_{\Omega_2}\|_2^2\).

In order to apply these three Lemmas, we split \(\|y_{\Omega_2} - P_{S_{\Omega_2}} y_{\Omega_2}\|_2^2\) into three terms and bound each with high probability.
\begin{equation}
\|y_{\Omega_2} - P_{S_{\Omega_2}} y_{\Omega_2}\|_2^2 = \|Y_{\Omega_2}\|_F^2 - y_{\Omega_2}^T U_{\Omega_2} (U_{\Omega_2}^T U_{\Omega_2})^{-1} U_{\Omega_2}^T y_{\Omega_2}.
\end{equation}
And write the second term of Equation (16) as
\begin{equation}
y_{\Omega_2}^T U_{\Omega_2} (U_{\Omega_2}^T U_{\Omega_2})^{-1} U_{\Omega_2}^T y_{\Omega_2} = \|W_{\Omega_2} U_{\Omega_2}^T U_{\Omega_2} y_{\Omega_2}\|_2^2,
\end{equation}
where \(W_{\Omega_2}^T W_{\Omega_2} = (U_{\Omega_2}^T U_{\Omega_2})^{-1}\). Under the assumptions of our theorems, \(U_{\Omega_2}^T U_{\Omega_2}\) is invertible, we have
\begin{equation}
\|W_{\Omega_2} U_{\Omega_2}^T U_{\Omega_2} y_{\Omega_2}\|_2^2 \leq \|W_{\Omega_2}\|_2^2 \|U_{\Omega_2}^T U_{\Omega_2} y_{\Omega_2}\|_2^2 = \|W_{\Omega_2}^T W_{\Omega_2}\|_2 \|U_{\Omega_2}^T y_{\Omega_2}\|_2^2 = \|(U_{\Omega_2}^T U_{\Omega_2})^{-1}\|_2 \|U_{\Omega_2}^T y_{\Omega_2}\|_2^2.
\end{equation}
Thus
\begin{equation}
\|y_{\Omega_2} - P_{S_{\Omega_2}} y_{\Omega_2}\|_2^2 \geq \|Y_{\Omega_2}\|_F^2 - \|(U_{\Omega_2}^T U_{\Omega_2})^{-1}\|_2 \|U_{\Omega_2}^T y_{\Omega_2}\|_2^2,
\end{equation}
and combining Lemma 4, Lemma 5, and Lemma 6, we have
\begin{equation}
\frac{m_2(1 - \alpha) - n_2 n_3 \mu(S)(1 - \gamma)}{n_1 n_3} \|Y\|_F^2 \leq \|y_{\Omega_2} - P_{S_{\Omega_2}} y_{\Omega_2}\|_2^2 \leq (1 + \alpha)\frac{m_2}{n_1 n_3} \|Y\|_F^2
\end{equation}
with probability at least \(1 - 4\delta\). \qed
IV. HYPOTHESIS TESTING

In this section, the hypothesis testing for the problem of matched subspace detection under tubal-sampling and elementwise-sampling will be given, respectively.

A. Hypothesis Testing under Tubal-sampling

The hypotheses can be represented as following:

$$
\begin{align*}
H_0 & : V \in S \\
H_1 & : V \notin S 
\end{align*}
$$

(17)

Under tubal-sampling, the test statistic is

$$
t(V_{\Omega_1}) = \|V_{\Omega_1} - P_{S_{\Omega_1}} * V_{\Omega_1}\|_F^2 \overset{H_1}{\gtrless} \frac{\eta}{H_0}.
$$

(18)

In the noiseless case, the detection threshold $\eta = 0$. Theorem 1 shows for $\delta > 0$ and $m_1 \geq \frac{2}{3} n_2 \mu(S) \ln(\frac{2n_2}{\delta})$

$$
\text{P}_{\text{D}} = \text{P}[t(V_{\Omega_1}) > 0|H_1] \geq 1 - 4\delta,
$$

and the probability of false alarm is

$$
\text{P}_{\text{FA}} = \text{P}[t(V_{\Omega_1}) > 0|H_0] = 0,
$$

since when $V \in S$, $\|V_{\Omega_1} - P_{S_{\Omega_1}} * V_{\Omega_1}\|_F^2 = 0$.

When there is Gaussian white noise $N \in R^{n_1 \times 1 \times n_3}$ with its entries $N(i, 1, k) \sim N(0, 1)$, $i \in [n_1]$, $j \in [n_3]$, the observed signal can be represented as $W = V_{\Omega_1} + N_{\Omega_1}$ where $N_{\Omega_1}$ obtained by the same sampling as $V_{\Omega_1}$. And the test statistic is represented as following

$$
t(W) = \|W - P_{S_{\Omega_1}} * W\|_F^2 \overset{H_1}{\gtrless} \frac{\eta_p}{H_0}.
$$

(19)

Thus, $t(W)$ is distributed as non-central $\chi^2$-distribution with degree of freedom $(m_1 - n_2)n_3$ and non-centrality parameter $\lambda_p^2 = \|V_{\Omega_1} - P_{S_{\Omega_1}} * V_{\Omega_1}\|_F^2$. Then for a requested probability of false alarm $P_{FA} = p$, the detection threshold $\eta_p$ can be obtained according to the following

$$
\text{P}[t(W) > \eta_p|H_0] \leq p,
$$

and it can be simplified as

$$
\text{P}[\chi^2_{(m_1 - n_2)n_3}(0) \leq \eta_p] \geq 1 - p.
$$

The detection probability $P_D = \text{P}[t(W) > \eta_p|H_1] = 1 - \text{P}[\chi^2_{(m_1 - n_2)n_3}(\lambda_p^2) \leq \eta_p]$, and the probability of false alarm $P_{FA} = p$. 

B. Hypothesis Testing under Elementwise-sampling

With the same hypotheses as above subsection, the test statistic is represented as following:

\[ t(V_\Omega^2) = \|v_\Omega^2 - P_{S_\Omega^2}v_\Omega^2\|^2_H \geq \eta. \]  

(20)

In the noiseless case, the detection threshold \( \eta = 0 \). Theorem 2 shows that for \( \delta > 0 \) and \( m_2 \geq \frac{8}{3}n_2n_3\mu(S)\ln(\frac{2n_2n_3}{\delta}) \), the probability of detection is \( P_D = \mathbb{P}[t(V_\Omega^2) > 0|H_1] \geq 1 - 4\delta \), and the probability of false alarm is \( P_{FA} = \mathbb{P}[t(V_\Omega^2) > 0|H_0] = 0 \), since when \( V \in S \), \( \|v_\Omega^2 - P_{S_\Omega^2}v_\Omega^2\|^2 = 0 \).

When there is Gaussian white noise \( N \in \mathbb{R}^{n_1 \times 1 \times n_3} \) with its entries \( N(i, 1, j) \sim \mathcal{N}(0, 1), \ i \in [n_1], j \in [n_3] \), the observed signal can be represented as \( W = V_\Omega^2 + N_\Omega^2 \) where \( N_\Omega^2 \) obtained by the same sampling as \( V_\Omega^2 \). Let \( w = \text{unfold}(W) \), and the test statistic can be represented as follows

\[ t(W) = \|w - P_{S_\Omega^2}w\|^2_H \geq \eta_p. \]  

(21)

Therefore, \( t(W) \) is distributed as a non-central \( \chi^2 \)-distribution with degree of freedom \( m_2 - n_2n_3 \) and non-centrality parameter \( \lambda_1^2 = \|v_\Omega^2 - P_{S_\Omega^2}v_\Omega^2\|^2_H \). Then for a requested probability of false alarm \( P_{FA} = p \), the detection threshold \( \eta_p \) can be obtained according to the following

\[ \mathbb{P}[t(W) > \eta_p|H_0] \leq p. \]

And the equation above can be rewritten as

\[ \mathbb{P}[\chi^2_{m_2 - n_2n_3}(0) \leq \eta_p] \geq 1 - p. \]

The detection probability \( P_D = \mathbb{P}[t(W) > \eta_p|H_1] = 1 - \mathbb{P}[\chi^2_{m_2 - n_2n_3}(\lambda_1^2) \leq \eta_p] \), and the probability of false alarm \( P_{FA} = p \).

V. TENSOR SUBSPACE DETECTION IN ARBITRARY SPACES

In former sections, the problem of tensor matched subspace detection was discussed. Motivated by [12] that tensor-tensor products based on linear transformation can be defined, we hope to find out the similar results to Theorem 1 and theorem 2. We first introduce some definitions of tensor-tensor product, and then we will give two theorems of tensor subspace detection in arbitrary spaces.

A. Tensor-tensor Products based on Arbitrary Linear Transforms

For convenience of the following discussion, we first introduce the operations of vec(\cdot) and tube(\cdot). For a tubal scalar \( a \in \mathbb{K}_n \), we use vec(\( a \)) to denote the vector version of \( a \), i.e., vec(\( a \))\( _i = a^{(i)} \). And tube(\( \cdot \)) is the inverse operation to vec(\( \cdot \)), i.e., \( a = \text{tube}(\text{vec}(a)) \).
Definition 11. \cite{12} Let $L : \mathbb{K}_n \rightarrow \mathbb{K}_n$ be an invertible linear transform defined by $L(y)^{(i)} = (M \text{vec}(y))^i$ for invertible $n \times n$ matrix $M$. Likewise, $L : \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined by applying $L$ to each tube of an $n_1 \times n_2 \times n_3$ tensor.

Motivated by the definitions of the t-product in \cite{9} and the cosine transform product $*_{c}$ in \cite{12}, the tensor-tensor product $*_{L}$ based on linear transform can also be defined by using a special structure block matrix by the frontal slices of the tensors such that the block structure can be block-diagonalized by matrix built from 1D discrete linear transform.

Definition 12. For $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $B \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, the $*_{L}$ of $A$ and $B$ can be implemented as following:

$$A *_{L} B = \text{lten}(\text{lmat}(A) \text{lmat}(B)),$$

where $\text{lmat}(\cdot)$ denotes the special structured block matrix of a tensor, and $\text{lten}(\cdot)$ is the inverse of $\text{lmat}(\cdot)$.

For an $n_1 \times 1 \times n_3$ tensor $V = U *_{L} C$ where $U \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $C \in \mathbb{R}^{n_2 \times 1 \times n_3}$ is equivalent to $\text{unfold}(V) = \text{lmat}(U) \text{unfold}(C)$.

The definition of tensor-tensor product $*_{L}$ based on linear transform can also be given as following.

Definition 13 (Tensor-scalar product $*_{L}$). For $a, b \in \mathbb{K}_n$, the tensor product $*_{L} : \mathbb{K}_n \times \mathbb{K}_n \rightarrow \mathbb{K}_n$ is defined as following.

$$L(a *_{L} b) = L(a) \bullet L(b)$$

where $\bullet$ denotes element-wise product.

Since a third-order tensor can be viewed as an matrix whose elements are tubal scalars, the tensor-tensor product of two tensors can be defined as following.

Definition 14 (Tensor-tensor product \cite{12}). The tensor-tensor product $C = A *_{L} B$ of $A \in \mathbb{R}^{n_1 \times n_2 \times k}$ and $B \in \mathbb{R}^{n_2 \times n_4 \times k}$ is a tensor of size $n_1 \times n_3 \times k$. $C(i,j,:) = \sum_{s=0}^{n_2} A(i,s,:) *_{L} B(s,j,:)$, for $i \in [n_1]$ and $j \in [n_3]$.

Let $L(A)$ denote the block-diagonal matrix of the transform domain tensor $L(A)$, i.e.,

$$L(A) \triangleq \begin{bmatrix}
L(A)^{(1)} \\
L(A)^{(2)} \\
\vdots \\
L(A)^{(n_3)}
\end{bmatrix},$$

where $L(A)^{(i)}$ is the block diagonal of $L(A)$.
and for the tensor-tensor product of $A \in \mathbb{R}^{n_1 \times n_2 \times k}$ and $B \in \mathbb{R}^{n_2 \times n_3 \times k}$, we have

$$L(A \ast_L B) = L(A) \ast L(B).$$

**Remark 1.** For a tubal scalar $a \in \mathbb{R}^{1 \times 1 \times n}$ or a tensor $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have $c\|a\|_2^2 = \|L(a)\|_2^2$ and $c\|A\|_2^2 = \|L(A)\|_2^2$, where $c$ is a constant.

**Definition 15.** Let $L(J) \in \mathbb{R}^{m \times m \times n}$ be so that $L(J)^{(i)} = J_{m \times m}$ for $i = 1, \ldots, n$. Then $J = L^{-1}(L(J))$ is the identity tensor under $\ast_L$ where $L^{-1}$ is the inverse of $L$.

**Definition 16.** A tensor $A \in \mathbb{R}^{m \times m \times n}$ is invertible under $\ast_L$ if there exists a tensor $A^{-1} \in \mathbb{R}^{m \times m \times n}$ for which $A \ast L A^{-1} = A^{-1} \ast L A = J$.

**Definition 17.** $A$ is $\ast_L$-orthogonal if $A^\dagger \ast_L A = A \ast_L A^\dagger = J$.

**Definition 18 (L-SVD).** The $L$-SVD of $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is given by $A = U \ast_L \Sigma \ast_L V^\dagger$, where $U$ and $V$ are orthogonal tensors of size $n_1 \times n_1 \times n_3$ and $n_2 \times n_2 \times n_3$ respectively, and $\Sigma$ is a diagonal tensor of size $n_1 \times n_2 \times n_3$. The entries in $\Sigma$ are called the singular values of $A$.

**Definition 19 (L-rank).** The tensor $L$-rank of $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as the number of non-zero tubal-scalars of $\Sigma$ in the $L$-SVD.

**Definition 20 (Tensor column subspace under $\ast_L$).** If $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $L$-rank of $r$, the $r$-dimensional tensor-column subspace $S_L$ spanned by columns of $A$ is defined as

$$S_L = \{X | X = A_1 \ast_L c_1 + A_2 \ast_L c_2 + \cdots + A_{n_2} \ast_L c_{n_2}\}$$

where $c_j, j \in [n_2]$, is an arbitrary tubal scalar of length $n_3$.

**Proposition 1.** If $S_L$ is spanned by columns of $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $P_L \triangleq A \ast_L (A^\dagger \ast_L A)^{-1} \ast_L A^\dagger$ is an orthogonal projection onto $S_L$ when $A^\dagger \ast_L A$ is invertible.

**Definition 21.** The coherence of an $r$-dimensional subspace $S_L$ is defined as

$$\mu(S_L) \triangleq \frac{1}{r} \max \|L(P_L)E_j\|_F^2,$$
B. Main Theorem under Tubal-sampling

Let \( V \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) denote a signal, and the tensor column subspace \( S_L \) be spanned by \( U \in \mathbb{R}^{n_1 \times n_2 \times n_3} \), so that \( \mathcal{P}_L = U \ast_L (U^\ast \ast_L U)^{-1} \ast_L U^\dagger \) is the orthogonal projection onto \( S_L \). Assume the dimension of \( S_L \) is \( n_2 \), and let \( V = X + \mathcal{Y} \) with its entries sampled uniformly with replacement where \( X \in S_L \) and \( \mathcal{Y} \in S_L^\perp \). With the same definition as section III-B, we define \( \mathcal{P}_{L_\Omega_1} = U_{\Omega_1} \ast_L (U_{\Omega_1}^\ast \ast_L U_{\Omega_1})^{-1} \ast_L U_{\Omega_1}^\dagger \) as the projection, and the main theorem of tensor subspace detection under tubal-sampling is given as following.

**Theorem 3.** Let \( \delta > 0 \) and \( m_1 \geq \frac{8}{3} n_2 \mu(S_L) \ln\left(\frac{2m_1}{\delta}\right) \). Then with probability at least \( 1 - 4\delta \),

\[
\frac{m_1 (1 - \alpha) - cn_2 \mu(S_L) \beta}{n_1} \| \mathcal{V} - \mathcal{P}_L \ast_L \mathcal{V} \|_F^2 \leq \| \mathcal{V}_{\Omega_1} - \mathcal{P}_{L_{\Omega_1}} \ast_L \mathcal{V}_{\Omega_1} \|_F^2 \leq (1 + \alpha) \frac{m_1}{n_1} \| \mathcal{V} - \mathcal{P}_L \ast_L \mathcal{V} \|_F^2
\]

(22)

holds, where \( \alpha = \sqrt{2 \left( \frac{n_1 \| \mathcal{Y} \|_F^2}{m_1 \| \mathcal{Y} \|_F^2} \right) \ln \left( \frac{1}{\delta} \right) + \frac{2(n_1 \| \mathcal{Y} \|_F^2 - \| \mathcal{Y} \|_F^2)}{3m_1 \| \mathcal{Y} \|_F^2} \ln \left( \frac{1}{\delta} \right)} \), \( \beta = \left( 1 + \sqrt{\frac{2}{3} \ln \left( \frac{1}{\delta} \right)} \right)^2 \), and \( \gamma = \sqrt{\frac{2n_2 \mu(S_L)}{3m_1} \ln \left( \frac{2m_1}{\delta} \right)} \).

**Proof:** In order to prove Theorem 3, we first give another two Lemmas and their proofs can be found in Appendix.

**Lemma 7.** With the same notations as Theorem 3,

\[
\left\| U_{\Omega_1}^\dagger \ast_L \mathcal{Y}_{\Omega_1} \right\|_F^2 \leq \frac{\beta m_1 n_2 \mu(S_L)}{n_1^2} \| \mathcal{Y} \|_F^2
\]

(23)

holds with probability at least \( 1 - \delta \).

**Lemma 8.** With the same notations as Theorem 3,

\[
\left\| \left( L(U_{\Omega_1})^T L(U_{\Omega_1}) \right)^{-1} \right\|_2 \leq \frac{n_1}{(1 - \gamma)m_1}
\]

(24)

holds with probability at least \( 1 - \delta \), and \( \gamma < 1 \).

Consider \( \| \mathcal{V}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{V}_{\Omega_1} \|_F^2 = \| \mathcal{Y}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{Y}_{\Omega_1} \|_F^2 \), then we split \( \| \mathcal{Y}_{\Omega_1} - \mathcal{P}_{S_{\Omega_1}} \ast \mathcal{Y}_{\Omega_1} \|_F^2 \) into three
terms in order to apply the two lemmas above.

\[
\|\mathcal{Y}_{\Omega_1} - \mathcal{P}_{L_{\Omega_1}} * L \mathcal{Y}_{\Omega_1}\|_F^2 = \frac{1}{c} \left\| L(\mathcal{Y}_{\Omega_1}) - L(\mathcal{P}_{L_{\Omega_1}}) L(\mathcal{Y}_{\Omega_1}) \right\|_F^2
\]

\[
= \frac{1}{c} \text{trace} \left( \left( L(\mathcal{Y}_{\Omega_1}) - L(\mathcal{P}_{L_{\Omega_1}}) L(\mathcal{Y}_{\Omega_1}) \right)^T \left( L(\mathcal{Y}_{\Omega_1}) - L(\mathcal{P}_{L_{\Omega_1}}) L(\mathcal{Y}_{\Omega_1}) \right) \right)
\]

\[
= \frac{1}{c} \text{trace} \left( L(\mathcal{Y}_{\Omega_1})^T L(\mathcal{Y}_{\Omega_1}) - L(\mathcal{Y}_{\Omega_1})^T L(\mathcal{P}_{L_{\Omega_1}}) L(\mathcal{Y}_{\Omega_1}) \right)
\]

\[
\geq \|\mathcal{Y}_{\Omega_1}\|_F^2 - \left\| L(U_{\Omega_1}^*) L(U_{\Omega_1}) \right\|_F^{-1}\left\| U_{\Omega_1}^* \mathcal{Y}_{\Omega_1}\|_F^2. \right. \]

(25)

and combining Lemma 1, Lemma 7, and Lemma 8, we have

\[
m_1(1 - \alpha) - c n_2 \mu(S_L) \frac{\beta}{(1 - \gamma)} \|\mathcal{Y}\|_F^2 \leq \|\mathcal{Y}_{\Omega_1} - \mathcal{P}_{L_{\Omega_1}} * L \mathcal{Y}_{\Omega_1}\|_F^2 \leq (1 + \alpha) m_1 \frac{\|\mathcal{Y}\|_F^2}{n_1} \]

(26)

with probability at least $1 - 4\delta$.

C. Main Theorem under Elementwise-sampling

Let $\mathcal{V} \in \mathbb{R}^{n_1 \times 1 \times n_3}$ denote a signal, and the tensor column subspace $S_L$ be spanned by $U \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, so that $\mathcal{P}_L = U * L (U^T * L U)^{-1} * L U^T$ is the orthogonal projection onto $S_L$. Assume the dimension of $S_L$ is $n_2$, and let $\mathcal{V} = \mathcal{X} + \mathcal{Y}$ with its entries sampled uniformly with replacement where $\mathcal{X} \in S_L$ and $\mathcal{Y} \in S_L^\perp$. Let $v_{\Omega_2} = \text{unfold}(\mathcal{Y}_{\Omega_2})$ and $P_{L_{\Omega_2}} = U_{\Omega_2} (U_{\Omega_2}^T U_{\Omega_2})^{-1} U_{\Omega_2}^T$, where $U_{\Omega_2}$ satisfies

\[
U_{\Omega_2}((j-1)n_1 + i,:) = \begin{cases} \text{imat}(U)((j-1)n_1 + i, :), & (i,j) \in \Omega_2; \\ 0, & \text{otherwise.} \end{cases} \]

(27)

If $\mathcal{Y} \in S$, then $\|\mathcal{V} - \mathcal{P}_L * L \mathcal{V}\|_F^2 = 0$ and $\|v_{\Omega_2} - P_{L_{\Omega_2}} v_{\Omega_2}\|_F^2 = 0$. Thus we have the following.

**Theorem 4.** Let $\delta > 0$, $m_2 \geq \frac{8}{3} n_2 n_3 b \mu(S_L) \ln(\frac{2m_2 \mu(S_L)}{\delta})$, and $b$ be a constant determined by $\text{imat}(U)$. Then with probability at least $1 - 4\delta$

\[
m_2(1 - \alpha) - n_2 n_3 \mu(S_L) \frac{\beta}{(1 - \gamma)} \|\mathcal{V} - \mathcal{P}_L * L \mathcal{V}\|_F^2 \leq \|v_{\Omega_2} - P_{L_{\Omega_2}} v_{\Omega_2}\|_F^2 \leq (1 + \alpha) m_2 \frac{n_1 n_3}{n_1 n_3} \|\mathcal{Y} - \mathcal{P}_L * L \mathcal{V}\|_F^2 \]

(28)

holds, where $\alpha = \sqrt{\frac{2(n_1 n_3) \|\mathcal{V}\|_F^2 - \|\mathcal{V}\|_F^2}{m_2} \ln(\frac{8}{\alpha})} + \frac{2(n_1 n_3) \|\mathcal{V}\|_F^2 - \|\mathcal{V}\|_F^2}{3m_2} \|\mathcal{Y}\|_F^2 \ln(\frac{1}{8})}$, $\beta = \left(1 + 2 \sqrt{\ln(\frac{1}{8})}\right)^2$, and $\gamma = \sqrt{\frac{8n_2 n_3 b \mu(S_L)}{3m_2} \ln(\frac{2m_2 \mu(S_L)}{\delta})}$.

**Proof:** In order to prove Theorem 4, we first give another two Lemmas and their proofs can be found in Appendix.
Lemma 9. With the same notations as Theorem 4,
\[
\|U_{\Omega_2}^T y_{\Omega_2}\|_2^2 \leq \beta \frac{m_2 n_2 b}{n_1 n_3} \mu(S_L) \|Y\|_F^2.
\] (29)
holds with probability at least 1 – δ.

Lemma 10. With the same notations as Theorem 4,
\[
\left\| (U_{\Omega_2}^T U_{\Omega_2})^{-1} \right\|_2 \leq \frac{n_1 n_3}{(1 - \gamma)m_2}
\] (30)
holds with probability at least 1 – δ, and γ < 1.

For \(\|v_{\Omega_2} - P_{L_{\Omega_2}} v_{\Omega_2}\|_2 = \|y_{\Omega_2} - P_{L_{\Omega_2}} y_{\Omega_2}\|_2^2\) and motivated by proof of theorem 1, we split \(\|y_{\Omega_2} - P_{L_{\Omega_2}} y_{\Omega_2}\|_2\) into three terms and bound each with high probability.
\[
\|y_{\Omega_2} - P_{L_{\Omega_2}} y_{\Omega_2}\|_2 = \|Y_{\Omega_2}\|_F - y_{\Omega_2}^T U_{\Omega_2} (U_{\Omega_2}^T U_{\Omega_2})^{-1} U_{\Omega_2}^T y_{\Omega_2}
\geq \|Y_{\Omega_2}\|_F - \left\| (U_{\Omega_2}^T U_{\Omega_2})^{-1} \right\|_2 \left\| U_{\Omega_2}^T y_{\Omega_2}\right\|_2.
\]

Combining Lemma 4, Lemma 9, and Lemma 10, we have
\[
\frac{m_2 (1 - \alpha) - n_2 n_3 \mu(S_L) \beta}{n_1 n_3} \|Y\|_F^2 \leq \|y_{\Omega_2} - P_{L_{\Omega_2}} y_{\Omega_2}\|_2 \leq (1 + \alpha) \frac{m_2}{n_1 n_3} \|Y\|_F^2
\]
with probability at least 1 – 4δ.

APPENDIX

PROOF OF LEMMAS

The following three versions of Bernstein’s inequalities are needed in our proofs.

Lemma 11 (Scalar Version). Let \(X_1, \ldots, X_m\) be independent zero-mean scalar variables. Suppose \(\rho_k = \mathbb{E}[X_k^2]\) and \(|X_k| \leq M\) almost surely for all \(k\). Then for any \(\tau > 0\),
\[
\mathbb{P} \left[ \sum_{k=1}^m X_k \geq \tau \right] \leq \exp \left( \frac{-\tau^2/2}{\sum_{k=1}^m \rho_k^2 + M\tau/3} \right).
\] (31)

Lemma 12 (Vector Version). Let \(X_1, \ldots, X_m\) be independent zero-mean random vectors with \(\sum_{k=1}^m \mathbb{E}\|X_k\|_2^2 \leq \Gamma\). Then for any \(\tau \leq \Gamma (\max_i \|X_i\|_2)^{-1}\),
\[
\mathbb{P} \left[ \left\| \sum_{k=1}^m X_k \right\|_2 \geq \sqrt{\Gamma + \tau} \right] \leq \exp \left( \frac{-\tau^2}{4\Gamma} \right).
\] (32)
Lemma 13 (Matrix Version). Let $X_1, \ldots, X_m$ be independent zero-mean square $r \times r$ random matrices.

Suppose $\rho_k = \max\{\|E[X_kX_k^T]\|_2, \|E[X_k^TX_k]\|_2\}$ and $\|X_k\|_2 \leq M$ almost surely for all $k$. Then for any $\tau > 0$,

$$\Pr\left[\left\| \sum_{k=1}^m X_k \right\|_2 > \tau \right] \leq 2\tau \exp\left(\frac{-\tau^2/2}{\sum_{k=1}^m \rho_k^2 + M\tau/3}\right).$$

Proof of Lemma 1: We use scalar version of Bernstein’s inequality to prove Lemma 1. Let $X_i = \|\Omega_i(i,:,:)\|_2^2 - E\left[\|\Omega_i(i,:,:)\|_2^2\right] = \|\Omega_i(i,:,:)\|_2^2 - \frac{1}{n_1} \|\Omega\|_F^4$, and $E[X_i] = 0$. In order to apply Bernstein’s inequality, we calculate $\rho_i^2$ and $M$. For

$$E[X_i^2] = E\left[\left(\|\Omega_i(i,:,:)\|_2^2 - \frac{1}{n_1} \|\Omega\|_F^4\right)^2\right]$$

$$= E\left[\sum_{j=1}^{n_1} \|\Omega(j,:,:)\|_2^4 \mathbf{1}_{(\Omega_i(i,j) = j)}\right] - \left(\frac{1}{n_1}\right)^2 \|\Omega\|_F^4$$

$$= \frac{1}{n_1} \sum_{j=1}^{n_1} \|\Omega(j,:,:)\|_2^4 - \left(\frac{1}{n_1}\right)^2 \|\Omega\|_F^4$$

$$\leq \frac{1}{n_1} \|\Omega\|_\infty^2 \|\Omega\|_F^4 - \left(\frac{1}{n_1}\right)^2 \|\Omega\|_F^4$$

and

$$|X_i| \leq \|\Omega\|_\infty^2 - \frac{1}{n_1} \|\Omega\|_F^4,$$

we set $\rho_i^2 = \frac{1}{n_1} \|\Omega\|_\infty^2 \|\Omega\|_F^4 - \left(\frac{1}{n_1}\right)^2 \|\Omega\|_F^4$ and $M = \|\Omega\|_\infty^2 - \frac{1}{n_1} \|\Omega\|_F^4$. Now we apply Bernstein’s inequality:

$$\Pr\left[\left| \sum_{i \in \Omega_i} X_i \right| > \tau \right] \leq 2\exp\left(\frac{-\tau^2/2}{\sum_{i \in \Omega_i} \rho_i^2 + M\tau/3}\right)$$

Let $\tau = \alpha \frac{n_1}{n_1} \|\Omega\|_F^2$ where $\alpha$ is defined in Theorem 1, and we have

$$(1 - \alpha) \frac{n_1}{n_1} \|\Omega\|_F^2 \leq \|\Omega_i\|_F^2 \leq (1 + \alpha) \frac{n_1}{n_1} \|\Omega\|_F^2$$

with the probability at least $1 - 2\delta$. \hfill \blacksquare

Proof of Lemma 2: Vector version of Bernstein’s inequality is used to prove Lemma 2. Let $X_i = \mathcal{U}(\mathcal{H}, \Omega_1(i,:), \cdot) \ast \mathcal{Y}((\Omega_1(i), 1,:))$. Since $\mathcal{Y} \in \mathcal{S}^\perp$, we have the followings.

$$E[X_i] = E\left[\sum_{j=1}^{n_1} \mathcal{U}(\mathcal{H}, j,:) \ast \mathcal{Y}(j, 1,:) \mathbf{1}_{(\Omega_1(i,j) = j)}\right]$$

$$= \frac{1}{n_1} \sum_{j=1}^{n_1} \mathcal{U}(\mathcal{H}, j,:) \ast \mathcal{Y}(j, 1,:) = 0,$$
\[
\sum_{j=1}^{m_1} E \|X_i\|_2^2 = \frac{m_1}{n_1} \sum_{j=1}^{n_1} \left\| \mathcal{U}^i(:, j, :) \ast \mathcal{Y}(j, 1, :) \right\|_F^2
\]

\[
= \frac{m_1}{n_1} \sum_{j=1}^{n_1} \sum_{k=1}^{n_3} \frac{1}{n_3} \left\| \tilde{\mathcal{U}}^i(:, j, k) \tilde{\mathcal{Y}}(j, 1, k) \right\|_2^2
\]

\[
\leq \frac{m_1}{n_1} \left( \frac{1}{n_3} \sum_{k=1}^{n_3} \left\| \tilde{\mathcal{U}}^i(:, j, k) \right\|_F \right) \left( \sum_{j=1}^{n_1} \sum_{k=1}^{n_3} \left\| \tilde{\mathcal{Y}}(j, 1, k) \right\|_2^2 \right)
\]

\[
\leq \frac{m_1}{n_1} \frac{n_2 \mu(S)}{n_1^2} \|\mathcal{Y}\|_F^2
\]

In order to apply Bernstein’s inequality, we set \( \Gamma = \frac{m_1 n_2 \mu(S)}{n_1^2} \|\mathcal{Y}\|_F^2 \) and \( \tau = \sqrt{4 \Gamma \ln(1/\delta)} \). When \( m_1 \geq \frac{4 n_1 \|\mathcal{Y}\|_F^2}{\|\mathcal{Y}\|_F^2} \cdot \ln \left( \frac{1}{\delta} \right) \), \( \sqrt{4 \Gamma \ln(1/\delta)} \leq \Gamma \left( \max_i \|X_i\|_2 \right)^{-1} \), and we have

\[
\left\| \mathcal{U}^i \ast \mathcal{Y}_{\Omega_i} \right\|_F^2 \leq \left( \sqrt{\Gamma} + \sqrt{4 \Gamma \ln(1/\delta)} \right)^2
\]

\[
= \beta \frac{m_1 n_2 \mu(S)}{n_1^2} \|\mathcal{Y}\|_F^2
\]

with the probability at least \( 1 - \delta \) where \( \beta \) is defined in Theorem 1.

**Proof of Lemma 3:** We use the matrix version of Bernstein’s inequality to calculate the bound of \( \|\mathcal{U}^i_{\Omega_i} \mathcal{U}^i_{\Omega_i}^{-1}\|_2 \). Let \( \mathcal{X}_k = \mathcal{U}^i_{\Omega_i(k)} \ast \mathcal{U}_{\Omega_i(k)} - \frac{1}{n_1} \mathcal{I}_{n_2} \) where the notation \( \mathcal{U}_{\Omega_i(k)} \) is the \( \Omega_i(k) \)th row of \( \mathcal{U} \), and \( \mathcal{I}_{n_2} \) is the \( n_2 \times n_2 \times n_3 \) identity tensor. Note that the random variable \( \mathcal{X}_k \) is zero mean, and \( \bar{\mathcal{X}}_k = \mathcal{U}^T_{\Omega_i(k)} \mathcal{U}_{\Omega_i(k)} - \frac{1}{n_1} \mathcal{I}_{n_2} \) is also zero mean where . For ease of notation we will denote \( \mathcal{U}_{\Omega_i(k)} \) as \( \mathcal{U}_k \) and \( \overline{\mathcal{U}}_{\Omega_i(k)} \) as \( \overline{\mathcal{U}}_k \). Since \( \|\mathcal{U}_k\|_F^2 = \|\mathcal{U}^i \ast \hat{\mathcal{E}}_k\|_F^2 = \|\mathcal{P}_S \ast \hat{\mathcal{E}}_k\|_F^2 \leq n_2 \mu(S)/n_1 \), we have

\[
\left\| \mathcal{U}^T_k \overline{\mathcal{U}}_k - \frac{1}{n_1} \mathcal{I}_{n_2} \right\|_2 \leq \max \left\{ \frac{n_2 \mu(S)}{n_1}, \frac{1}{n_1} \right\}.
\]
And let $M = \frac{n_2 \mu(S)}{n_1}$. Next, we calculate $\|E[\overline{X}_k \overline{X}_k^T]\|_2$ and $\|E[\overline{X}_k^T \overline{X}_k]\|_2$.

\[
\begin{align*}
\|E[\overline{X}_k \overline{X}_k^T]\|_2 &= \|E[\overline{X}_k^T \overline{X}_k]\|_2 \\
&= \left\|E\left[\overline{U}_k^T \overline{U}_k - \frac{1}{n_1} I_{n_2}\right]\right\|_2 \\
&= \left\|E\left[\overline{U}_k^T \overline{U}_k \overline{U}_k^T \overline{U}_k - \frac{2}{n_1} \overline{U}_k^T \overline{U}_k + \frac{1}{n_1^2} I_{n_2}\right]\right\|_2 \\
&= \left\|E\left[\overline{U}_k^T \overline{U}_k \overline{U}_k^T \overline{U}_k - \frac{1}{n_1^2} I_{n_2}\right]\right\|_2 \\
&\leq \max\left\{\|E[\overline{U}_k^T \overline{U}_k]\|_2, \frac{1}{n_1^2}\right\} \\
&\leq \max\left\{\frac{n_2 \mu(S)}{n_1} \|E[\overline{U}_k]\|_2, \frac{1}{n_1^2}\right\} \\
&= \max\left\{\frac{n_2 \mu(S)}{n_1} \|I_{n_2}\|_2, \frac{1}{n_1^2}\right\} \\
&= \frac{n_2 \mu(S)}{n_1^2}.
\end{align*}
\]

Thus let $\rho^2 = \frac{n_2 \mu(S)}{n_1^2}$. By Bernstein Inequality, we have

\[
P\left[\left\|\sum_{k \in \Omega_1} \left(\overline{U}_k^T \overline{U}_k - \frac{1}{n_1} I_{n_2}\right)\right\|_2 > \tau\right] = P\left[\left\|\sum_{k \in \Omega_1} \left(\overline{U}_k^T \overline{U}_k - \frac{1}{n_1} I_{n_2}\right)\right\|_2 > \tau\right] \\
\leq 2n_2 n_3 \exp\left(\frac{-\tau^2/2}{m_1 \rho^2 + M \tau / 3}\right).
\]

We restrict $\tau$ to be $M \tau \leq m_1 \rho^2$, so the equation can be simplified as

\[
P\left[\left\|\sum_{k \in \Omega_1} \left(\overline{U}_k^T \overline{U}_k - \frac{1}{n_1} I_{n_2}\right)\right\|_2 > \tau\right] \leq 2n_2 n_3 \exp\left(\frac{-3n_1^2 \tau^2}{8m_1 n_2 \mu(S)}\right).
\]

Now set $\tau = \gamma m_1 / n_1$ with $\gamma$ defined in the statement of Theorem 1. We assume that $\gamma < 1$, and $M \tau \leq m_1 \rho^2$ holds. Then we have

\[
P\left[\left\|\sum_{k \in \Omega_1} \left(\overline{U}_k^T \overline{U}_k - \frac{1}{n_1} I_{n_2}\right)\right\|_2 \leq \frac{m_1}{n_1} \gamma\right] \geq 1 - \delta.
\]

We note that $P\left[\left\|\sum_{k \in \Omega_1} \left(\overline{U}_k^T \overline{U}_k - \frac{1}{n_1} I_{n_2}\right)\right\|_2 \leq \frac{m_1}{n_1} \gamma\right]$ implies that the minimum singular value of $\sum_{k \in \Omega_1} \left(\overline{U}_k^T \overline{U}_k\right)$ is at least $(1 - \gamma) \frac{m_1}{n_1}$. This in turn implies that

\[
\left\|\left(\sum_{k \in \Omega_1} \overline{U}_k^T \overline{U}_k\right)^{-1}\right\|_2 \leq \frac{n_1}{(1 - \gamma)m_1}.
\]
That means that
\[
\left\| \left( \mathcal{U}_{\Omega_1}^\dagger \ast \mathcal{U}_{\Omega_1} \right)^{-1} \right\|_2 \leq \frac{n_1}{(1 - \gamma)m_1}
\]
holds with the probability at least \(1 - \delta\).

**Proof of Lemma 4**: We use scalar version of Bernstein’s inequality to prove Lemma 4. Let \(X_{ij} = \mathcal{Y}_{\Omega_j}^2(i, 1, j) - \mathbb{E} \left[ \mathcal{Y}_{\Omega_j}^2(i, 1, j) \right] = \mathcal{Y}_{\Omega_j}^2(i, 1, j) - \frac{1}{n_1n_3} \| \mathcal{Y} \|^2_F\), \((i, j) \in \Omega_2\), and \(\mathbb{E} [X_{ij}] = 0\). In order to apply Bernstein inequality, we calculate \(\rho_{ij}^2\) and \(M\). For
\[
\mathbb{E} \left[ X_{ij}^2 \right] = \mathbb{E} \left[ \left( \mathcal{Y}_{\Omega_j}^2(i, 1, j) - \frac{1}{n_1n_3} \| \mathcal{Y} \|^2_F \right)^2 \right] \leq \frac{1}{n_1n_3} \| \mathcal{Y} \|^2_\infty \| \mathcal{Y} \|^2_F - \left( \frac{1}{n_1n_3} \right)^2 \| \mathcal{Y} \|^4_F,
\]
and
\[
|X_{ij}| \leq \| \mathcal{Y} \|^2_\infty - \frac{1}{n_1n_3} \| \mathcal{Y} \|^2_F,
\]
we set \(\rho_{ij}^2 = \frac{1}{n_1n_3} \| \mathcal{Y} \|^2_\infty \| \mathcal{Y} \|^2_F - \left( \frac{1}{n_1n_3} \right)^2 \| \mathcal{Y} \|^4_F\) and \(M = \| \mathcal{Y} \|^2_\infty - \frac{1}{n_1n_3} \| \mathcal{Y} \|^2_F\). Now we apply Bernstein Inequality:
\[
\mathbb{P} \left[ \sum_{(i,j) \in \Omega_2} X_{ij} > \tau \right] \leq 2 \exp \left( \frac{-\tau^2/2}{\sum_{(i,j) \in \Omega_2} \rho_{ij}^2 + M \tau/3} \right)
\]
Let \(\tau = \alpha \frac{m_2}{n_1n_3} \| \mathcal{Y} \|^2_F\), where \(\alpha\) is defined in Theorem 2, and we have
\[
(1 - \alpha) \frac{m_2}{n_1n_3} \| \mathcal{Y} \|^2_F \leq \| \mathcal{Y} \|^2_\infty \leq (1 + \alpha) \frac{m_2}{n_1n_3} \| \mathcal{Y} \|^2_F
\]
with the probability at least \(1 - 2\delta\).

**Proof of Lemma 5**: We use vector version of Bernstein’s inequality to prove Lemma 5. Let \(X_{ij} = U^T \cdot (j - 1)n_1 + i) y_{(j-1)n_1+i}, \{(i,j) \in \Omega_2\}\). Since \(\mathcal{Y} \in \mathcal{S}_1\), we have the followings.
\[
\mathbb{E} \left[ X_{ij} \right] = \mathbb{E} \left[ \sum_{k=1}^{n_1n_3} U^T(:, k) y_k \mathbb{I}_{\{(j-1)n_1+i=k\}} \right] = \frac{1}{n_1n_3} \sum_{k=1}^{n_1n_3} U^T(:, k) y_k = 0,
\]
\[
\sum_{(i,j) \in \Omega_2} \mathbb{E} \left[ \| X_{ij} \|^2_2 \right] = \frac{m_2}{n_1n_3} \sum_{k=1}^{n_1n_3} \| U^T(:, k) y_k \|^2_2
\]
\[
= \frac{m_2}{n_1n_3} \sum_{k=1}^{n_1n_3} \| U^T(:, k) \|^2 \| y_k \|^2_2
\]
\[
\leq \frac{m_2}{n_1n_3} \frac{n_2\mu(S)}{n_1} \| \mathcal{Y} \|^2_F.
\]
In order to apply Bernstein’s inequality, we set $\Gamma = \frac{m_2 - n_2\mu(S)}{n_1 n_2} \|Y\|_F^2$ and $\tau = \sqrt{4\Gamma \ln(1/\delta)}$. When $m_2 \geq 4n_1n_2\|Y\|_F^2 \ln \left(\frac{1}{\delta}\right)$, $\sqrt{4\Gamma \ln(1/\delta)} \leq \Gamma \left(\max_{ij} \|X_{ij}\|_2\right)$, and we have

$$\|U_{\Omega_2}y_{\Omega_2}\|_2^2 \leq \left(\sqrt{\Gamma} + \sqrt{4\Gamma \ln(1/\delta)}\right)^2 = \left(1 + 2\sqrt{\ln(1/\delta)}\right)^2 \frac{m_2n_2\mu(S)}{n_1^2n_3} \|Y\|_F^2.$$

with the probability at least $1 - \delta$ where $\beta$ is defined in Theorem 2.

Proof of Lemma 6: We use matrix version of Bernstein Inequality to calculate the bound of $\left\|(U_{\Omega_2}^T U_{\Omega_2})^{-1}\right\|_2$. Let $X_k = U_{\Omega_2}((j-1)n_1 + i) U_{\Omega_2}^T ((j-1)n_1 + i) - \frac{1}{n_1 n_3} I_{n_2 n_3}$ where $(i,j) \in \Omega_2$, $k = (j-1)n_1 + i$, $U_{\Omega_2}((j-1)n_1 + i)$ is the $k^{th}$ rows of $U_{\Omega_2}$, and $I_{n_2 n_3}$ is the identity matrix. Note that this random variable is zero mean. Next we compute $\rho_k^2$ and $M$. Since $\Omega_2$ is chosen uniformly with replacement, the $X_k$ are identically distributed, and $\rho$ does not depend on $k$. For ease of notation we will denote $U_{\Omega_2}((j-1)n_1 + i)$ as $U(k)$. Using the fact represented in [11] that if $A$ and $B$ are positive semi-definite matrices, then $\|A - B\|_2 \leq \max\{\|A\|_2,\|B\|_2\}$. Since $\|U(k)\|_2^2 = \|U^T e_k\|_2^2 = \|P_k^T e_k\|_2^2 \leq \frac{n_2\mu(S)}{n_1}$ where $e_k$ is a standard matrix basis, we have

$$\left\|U(k)U(k)^T - \frac{1}{n_1 n_3} I_{n_2 n_3}\right\|_2 \leq \max\left\{\frac{n_2\mu(S)}{n_1}, \frac{1}{n_1 n_3}\right\},$$

and we let $M = \frac{n_2\mu(S)}{n_1}$. For $\rho$ we note

$$\left\|E\left[X_k X_k^T\right]\right\|_2 = \left\|E\left[X_k X_k^T\right]\right\|_2 = \left\|E\left[U(k)U(k)^T - \frac{1}{n_1 n_3} I_{n_2 n_3}\right]^2\right\|_2 \leq \left\|E\left[U(k)U(k)^T U(k)U(k)^T\right] - \frac{1}{n_1^2 n_3^2} I_{n_2 n_3}\right\|_2 \leq \max\left\{\left\|E\left[U(k)U(k)^T U(k)U(k)^T\right]\right\|_2, \frac{1}{n_1^2 n_3^2}\right\} \leq \max\left\{\frac{n_2\mu(S)}{n_1}, \left\|E\left[U(k)U(k)^T\right]\right\|_2, \frac{1}{n_1^2 n_3^2}\right\} \leq \max\left\{\frac{n_2\mu(S)}{n_1}, \left\|I_{n_2 n_3}\right\|_2, \frac{1}{n_1^2 n_3^2}\right\} = \frac{n_2\mu(S)}{n_1 n_3},$$

Thus we let $\rho^2 = \frac{n_2\mu(S)}{n_1^2 n_3}$. Noncommutative Bernstein Inequality can be used now. For simplifying the
denominator of the exponent, we restrict \( \tau \) as \( M \tau \leq m \rho^2 \). Then we get

\[
2n_2n_3 \exp \left( \frac{-x^2}{2 (m_2 \rho^2 + M \tau)} \right) \leq 2n_2n_3 \exp \left( \frac{-x^2}{4m_2 \mu(S) n_1 n_3} \right),
\]

and thus

\[
P \left( \left\| \sum_{(i,j) \in \Omega_2} \left( U(k)U(k)^T - \frac{1}{n_1 n_3} I_{n_2 n_3} \right) \right\|_F \geq \tau \right) \leq 2n_2n_3 \exp \left( \frac{-3n_1^2 n_3 \tau^2}{8m_2 n_2 \mu(S)} \right).
\]

Set \( \tau = \frac{m_2}{n_1 n_3} \), where \( \gamma \) have been defined in the statement of Theorem 2. We assume \( \gamma < 1 \), and \( M \tau \leq m_2 \rho^2 \) holds. With \( k \) defined above, we have

\[
P \left( \left\| \sum_{(i,j) \in \Omega_2} \left( U(k)U(k)^T - \frac{1}{n_1 n_3} I_{n_2 n_3} \right) \right\|_F \leq \frac{m_2}{n_1 n_3} \gamma \right) \geq 1 - \delta.
\]

Since \( \left\| \sum_{(i,j) \in \Omega_2} (U(k)U(k)^T - \frac{1}{n_1 n_3} I_{n_2 n_3}) \right\|_F \leq \frac{m_2}{n_1 n_3} \gamma \), the minimum singular value of \( \sum_{(i,j) \in \Omega_2} U(k)U(k)^T \) is at least \( (1 - \gamma) \frac{m_2}{n_1 n_3} \). This implies that

\[
\left\| \left( \sum_{(i,j) \in \Omega_2} U(k)U(k)^T \right)^{-1} \right\|_2 \leq \frac{n_1 n_3}{(1 - \gamma)m_2}.
\]

That means that

\[
\left\| \left( U_{\Omega_2}^T U_{\Omega_2} \right)^{-1} \right\|_2 \leq \frac{n_1 n_3}{(1 - \gamma)m_2}
\]

holds with the probability at least \( 1 - \delta \). ■

**Proof of lemma 7:** Let \( X_i = \mathcal{U}(:, \Omega_1(i), :) * \mathcal{L} \mathcal{Y}(\Omega_1(i), 1,:) \). Since \( \mathcal{Y} \in S^+_L \), we have the followings.

\[
\mathbb{E} [X_i] = \mathbb{E} \left[ \sum_{j=1}^{n_1} \mathcal{U}(::, j :) * \mathcal{L} \mathcal{Y}(j, 1, :) \mathbb{I}_{\Omega_1(j)=j} \right] = \frac{1}{n_1} \sum_{j=1}^{n_1} \mathcal{U}(::, j :) * \mathcal{L} \mathcal{Y}(j, 1, :) = 0,
\]

\[
\sum_{j=1}^{m_1} \mathbb{E} \|X_i\|_2^2 = \frac{m_1}{n_1} \sum_{j=1}^{n_1} \left\| \mathcal{U}(::, j :) * \mathcal{L} \mathcal{Y}(j, 1, :) \right\|_F^2 \leq \frac{m_1 n_2 \mu(S)}{n_1} c \|\mathcal{Y}\|_F^2 \leq \frac{m_1 n_2 \mu(S)}{n_1} \frac{c}{n_1^2} \|\mathcal{Y}\|_F^2.
\]
In order to apply Bernstein’s inequality, we set \( \Gamma = \frac{m_1 n_2 \mu(S)}{n_1} \|Y\|_F^2 \) and \( \tau = 4 \sqrt{\log \frac{1}{\delta}} \). When \( m \geq 4 \frac{n_2 \|Y\|_F^2}{\|Y\|_F^2} \ln \left( \frac{1}{\delta} \right) \), \( \sqrt{\Gamma \ln(1/\delta)} \leq \Gamma \left( \max_i \|X_i\|_2 \right)^{-1} \), and we have

\[
\|U_{\Omega_i}^\top L Y_{\Omega_i}\|_F^2 \leq (\sqrt{\Gamma} + \sqrt{4 \Gamma \ln(1/\delta)})^2 = \beta \frac{m_1 n_2 \mu(S)}{n_1^2} \|Y\|_F^2
\]

with the probability at least \( 1 - \delta \) where \( \beta \) is defined in Theorem 3.

**Proof of Lemma 8:** We use the matrix version of Bernstein’s inequality to calculate the bound of \( \|L (U_{\Omega_1})^T L (U_{\Omega_1})^{-1}\|_2 \). Let \( X_k = \frac{L (U_{\Omega_1})^T L (U_{\Omega_1})}{n_1} I_{n_2 n_3} \) where the notation \( U_{\Omega_1(k)} \) is the \( \Omega_1(k) \)th row of \( U \), and the random variable \( X_k \) is zero mean. For ease of notation we will denote \( \frac{L (U_{\Omega_1})^T L (U_k)}{n_1} I_{n_2 n_3} \) as \( \tilde{L}(U_k) \). Since \( \|L(U_k)^T L(U_k)\|_F^2 = \|L(\tilde{L})E_j\|_F^2 = \|L(\tilde{L})E_j\|_F^2 \leq n_2 \mu(S_L)/n_1 \), we have

\[
\left\| \frac{L (U_k)^T L (U_k) - \frac{1}{n_1} I_{n_2 n_3}}{n_1} \right\|_2 \leq \max \left\{ \frac{n_2 \mu(S_L)}{n_1}, \frac{1}{n_1} \right\}.
\]

And let \( M = \frac{n_2 \mu(S)}{n_1} \). Next, we calculate \( \|E[X_k X_k^T]\|_2 \) and \( \|E[X_k^T X_k]\|_2 \).

\[
\|E[X_k X_k^T]\|_2 = \|E[X_k^T X_k]\|_2
\]

Thus let \( \rho^2 = \frac{n_2 \mu(S)}{n_1^2} \). By Bernstein Inequality, we have

\[
P \left( \left\| \sum_{k \in \Omega_1} \left( \frac{L (U_k)^T L (U_k) - \frac{1}{n_1} I_{n_2 n_3}}{n_1} \right) \right\|_2 > \tau \right) \leq 2 n_2 n_3 \exp \left( \frac{-\tau^2 / 2}{m_1 \rho^2 + M \tau / 3} \right).
\]

We restrict \( \tau \) to be \( M \tau \leq m_1 \rho^2 \), so the equation can be simplified as

\[
P \left( \left\| \sum_{k \in \Omega_1} \left( \frac{L (U_k)^T L (U_k) - \frac{1}{n_1} I_{n_2 n_3}}{n_1} \right) \right\|_2 > \tau \right) \leq 2 n_2 n_3 \exp \left( \frac{-3 n_1^2 \tau^2}{8 m_1 n_2 \mu(S)} \right).
\]
Now set $\tau = \gamma m_1 / n_1$ with $\gamma$ defined in the statement of Theorem 2. We assume that $\gamma < 1$, and $M \tau \leq m_1 p^2$ holds. Then we have

$$\mathbb{P} \left[ \left\| \sum_{k \in \Omega_1} \left( \frac{L(U_k)^T L(U_k)}{n_1} - \frac{1}{n_1} I_{n_2 n_3} \right) \right\|_2 \leq \frac{m_1}{n_1} \gamma \right] \geq 1 - \delta.$$  

We note that $\mathbb{P} \left[ \left\| \sum_{k \in \Omega_1} \left( \frac{L(U_k)^T L(U_k)}{n_1} - \frac{1}{n_1} I_{n_2 n_3} \right) \right\|_2 \leq \frac{m_1}{n_1} \gamma \right]$ implies that the minimum singular value of $\sum_{k \in \Omega_1} \left( \frac{L(U_k)^T L(U_k)}{n_1} \right)$ is at least $(1 - \gamma) \frac{m_1}{n_1}$. This in turn implies that

$$\left\| \left( \sum_{k \in \Omega_1} \frac{L(U_k)^T L(U_k)}{n_1} \right)^{-1} \right\|_2 \leq \frac{n_1}{(1 - \gamma) m_1}.$$  

That means that

$$\left\| \left( \frac{L(U_{\Omega_1})^T L(U_{\Omega_1})}{n_1} \right)^{-1} \right\|_2 \leq \frac{n_1}{(1 - \gamma) m_1}$$  

holds with the probability at least $1 - \delta$.  

**Proofs of Lemma 9 and Lemma 10:** The proofs of Lemma 9 and Lemma 10 are consistent with the proofs of Lemma 5 and Lemma 6. According to the definition of $U_{\Omega_2}$, we have the following:

$$\| U_{\Omega_2}((j - 1)n_1 + i,:) \|^2_2 \leq b \frac{n_2 \mu(S_L)}{N_1}, \quad (i, j) \in \Omega_2, \quad (34)$$

where $b$ is a constant, i.e., $b = 1$ if the linear transformation $L$ is FFT. Substitute (34) into Proof of Lemma 5 and Proof of Lemma 6, we obtain the proof of Lemma 9 and Lemma 10.  

**REFERENCES**

[1] Yongjian Liu, Guoming Gao, and Yinafeng Gu, “Tensor matched subspace detector for hyperspectral target detection,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 1967–1974, December 2017.

[2] Yeqing Li, Chen Chen, and Junzhou Huang, “Transformation-invariant collaborative sub-representation,” in *IEEE 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 3738–3743.

[3] Hadi Sariedeen, Mohammad M. Mansour, and Ali Chehab, “Efficient subspace detection for high-order mimo systems,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 1001–1005.

[4] Mohammad M. Mansour, “A near-ml mimo subspace detection algorithm,” *IEEE Signal Processing Letters*, vol. 22, pp. 408–412, April 2015.
[5] Jun Wang and Kai Xu, “Shape detection from raw lidar data with subspace modeling,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 9, pp. 2137–2150, August 2017.

[6] Laura Balzano, Benjamin Recent, and Robert Nowak, “High-dimensional matched subspace detection when data are missing,” *IEEE International Symposium on Information Theory*, pp. 1638–1642, June 2010.

[7] Dejiao Zhang and Laura Balzano, “Matched subspace detection using compressively sampled data,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 4601–4605.

[8] Xiao-Yang Liu, Shuchin Aeron, Vaneet Aggarwal, and Xiaodong Wang, “Low-tubal-rank tensor completion using alternating minimization,” in *arXiv preprint arXiv:1610.01690*, 2016, pp. 1–48.

[9] M.E. Kilmer, K. Braman, N. Hao, and R.C. Hoover, “Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging,” *SIAM Journal on Matrix Analysis and Applications*, vol. 34, pp. 148–172, February 2013.

[10] KILMER M.E and C.D. Martin, “Factorization strategies for third-order tensors,” *Linear Algebra and its Applications*, vol. 435, pp. 641–658, August 2011.

[11] Zemin Zhang and Shuchin Aeron, “Exact tensor completion using t-svd,” *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1511–1526, March 2017.

[12] E. Kernfeld, M. Kilmer, and S. Aeron, “Tensor-tensor products with invertible linear transforms,” *Linear Algebra and its Applications*, vol. 485, pp. 545–570, November 2015.