Compressive tensor principal component pursuit using t-SVD

Haodong Sun[1], Tianwei Yue[2], Yao Wang[3,*], Shaobo Lin[4]

Abstract

This paper considers the problem of recovering a data tensor, which is the superposition of a low-rank component and a sparse component, from a small set of linear measurements. The interest in this problem is motivated by applications in compressed sensing of high-dimensional data with low-complexity structures. Our model is based on tensor-Singular Value Decomposition (t-SVD) and its induced tensor nuclear norm. In fact, we can derive a new notion of tensor rank by this factorization, denoted as tensor tubal rank. In this work, we prove that under suitable assumptions we can recover each component of the target tensor exactly by solving a convex problem with overwhelming probability. In the end, we show the result of numerical experiments conducted on several real data sets to illustrate the merit of our model.

1. Introduction

RECENTLY there has been a tremendous amount of interest in the problem of exploiting low-complexity structure in data lying in high-dimensional space, such as sparse signals or low-rank structures. Data in the form of multidimensional array is also referred as tensor. Naturally, we represent data in this way to avoid the loss of structure information. In this work, we mainly focus on the third order tensor. The huge progress made in compressed sensing has showed that it is possible to perfectly recover low-complexity signals from a small set of partially sampled measurements or corrupted observations. Generally, the tensor recovery is achieved by solving a optimal problem whose objective function is a weighted combination of several regularizations revealing the corresponding low-complexity properties. In fact, for exploiting low-rank structure, the strategy of recovery greatly relies on algebraic decomposition methods, such as CANDECOMP/PARAFAC(CP) decomposition[1] and Tucker decomposition[1]. Here we use t-SVD proposed in [2] due to its good properties.

Robust Principle Component Analysis (RPCA) proposed in [3] is an efficient algorithm with theoretical guarantee for exact recovery. Suppose that we have a data matrix $X \in \mathbb{R}^{n_1 \times n_2}$ which is the superposition of a low-rank component L_0 and a sparse component S_0. It has been proved in [3] that if L_0 meet some incoherent conditions and S_0 satisfies some probabilistic assumptions, one can exactly recover both components with high probability by solving a convex program called Principal Component Pursuit (PCP):

$$\min_{L,S} ||L||_* + \lambda ||S||_1, \quad s.t. \, X = L + S, \quad (1)$$

where $||L||_*$ denotes the nuclear norm and $||S||_1$ denotes the l-1 norm, $\lambda = 1/\sqrt{\max(n_1, n_2)}$. By the standard Alternating Direction Method of Multipliers (ADMM), the PCP can be solved easily in polynomial time. However, this method can only be used to manipulate two-dimensional data. To overcome this shortcoming, Tensor Robust Principle Component Analysis (TRPCA) was proposed in [4]. Moreover, in many scenarios it is meaningful to develop models recovering low-rank and sparse components from incomplete measurements. The compressive RPCA was studied in [5] for two-dimensional

data. Based on these works, it is interesting to consider the compressive RPCA problem for third order tensor.

In this paper, we study the Tensor Compressive Principal Component Pursuit (TCPCP). Consider we have a third order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ ($n_1 = n_2$) such that $\mathcal{X} = \mathcal{L}_0 + \mathcal{S}_0$ where $\mathcal{L}_0$ is the low-rank component and $\mathcal{S}_0$ is the sparse component. We prove that under some assumptions on $\mathcal{L}_0$, $\mathcal{S}_0$ and the dimension of sampling space, one can exactly recover both low-rank and sparse components from undersampled measurements by solving the following optimal problem:

$$\min_{\mathcal{L},\mathcal{S}} ||\mathcal{L}||_* + \lambda ||\mathcal{S}||_1, \quad s.t. \, \mathcal{P}_Q(\mathcal{L}_0 + \mathcal{S}_0) = \mathcal{P}_Q(\mathcal{L} + \mathcal{S}), \quad (2)$$

where Q is randomly chosen subspace of $\mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{P}_Q$ is the sampling operator, $\mathcal{L}_0, \mathcal{S}_0$ are low-rank and sparse components respectively. The tensor nuclear norm in this model is defined based on t-SVD factorization.

A. Related Work

1) Tensor Principle Component Analysis using t-SVD

Consider $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $n_1 \geq n_2$. Suppose that $\mathcal{X}$ can be decomposed as the sum of a low-rank tensor $\mathcal{L}_0$ and a sparse tensor $\mathcal{S}_0$. [4] tries to recover both components by solving the following program called Tensor PCP (TPCP):

$$\min_{\mathcal{L},\mathcal{S}} ||\mathcal{L}||_* + \lambda ||\mathcal{S}||_1, \quad s.t. \, \mathcal{X} = \mathcal{L} + \mathcal{S}, \quad (3)$$

where $\lambda = 1/\sqrt{n_1 n_3}$. The tensor nuclear norm here is induced by t-SVD.

In [4] the authors show that for μ-incoherent $\mathcal{L}_0$ with tubal rank on the order of $n_2 n_3 / \mu (\log n_1 n_3)^2$ and $\mathcal{S}_0$ whose number of nonzero entries is on the order of $n_1 n_2 n_3$, the exact recovery is guaranteed with overwhelming probability. Similar to [3], they make an assumption that the locations of nonzero entries of $\mathcal{S}_0$ are randomly distributed. Note that when n_3 is equal to 1, the TRPCA case reduces to RPCA case. So we can view TRPCA as an extension of RPCA. This property is closed related with t-SVD and we will introduce its conception and properties later.

2) Compressive Principle Component Pursuit

Let $L_0, S_0 \in \mathbb{R}^{n_1 \times n_2}$ with $n_1 \geq n_2$. Suppose L_0 is μ-incoherent and its rank is on the order of $n_2/\mu log^2 n_1$, and the sparsity of S_0 is on the order of $n_1 n_2$. Similarly we assume that the entries in support set of S_0 are randomly distributed.

[5] studies the following convex program called Compressive Principle Component Pursuit:

$$\min_{L,S} ||L||_* + \lambda ||S||_1, \quad s.t.\ D = P_Q(L+S), \qquad (4)$$

where Q is a subspace of $\mathbb{R}^{n_1 \times n_2}$, P_Q is the sampling operator, $\lambda = 1/\sqrt{n_1}$. With the assumptions above, [5] gives a tight bound on the number of measurements needed to recover (L_0, S_0) from D by solving CPCP. This bound is on the order of $(\rho n_1 n_2 + n_1 r) \cdot \log^2 n_1$, where ρ is the fraction of nonzero entries of S_0, r is the rank of L_0. The bound is closely related to the intrinsic degrees of freedom in (L_0, S_0).

B. Organization of the Paper

The rest of this paper is organized as follows. Section 2 introduces some notations and preliminaries. We show the main result in section 3 and delay the proof to section 4. In section 5 we report the results of several experiments. Finally we conclude the whole work and outline future research.

2. NOTATIONS AND PRELIMINARIES

Before the main part of this article, we firstly introduce the notations used throughout the paper and some basic definitions and theorems about tensor.

In this paper, we denote tensors by Euler script letter, e.g., $\mathcal{A}$. Matrices are denoted by capital letters, e.g., A. We denote vectors by boldface lowercase letters, e.g., $\mathbf{a}$. Scalars are denoted by lowercase letters, e.g., a. For a third order tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, its (i,j,k)-th entry is denoted as $\mathcal{A}_{ijk}$ or a_{ijk}. We use the MatLab notation $\mathcal{A}(i,:,:), \mathcal{A}(:,i,:), \mathcal{A}(:,:,i)$ to denote the i-th horizontal, lateral and frontal slice of tensor. Moreover, the frontal slice is also denoted as $A^{(i)}$. We denote $\mathcal{A}(i,j,:)$ as the tube of tensor. $\mathrm{Tr}(\cdot)$ denotes the matrix trace. The inner product of two matrices is $\langle A, B \rangle = Tr(A^*B)$. The inner product of $\mathcal{A}$ and $\mathcal{B}$ in $\mathbb{C}^{n_1 \times n_2 \times n_3}$ is $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_1^{n_3} \langle A^{(i)}, B^{(i)} \rangle$.

For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, with the matlab command `fft`, we use $\bar{\mathcal{A}}$ to denote the result of discrete Fourier transform of $\mathcal{A}$ along the third dimension, i.e., $\bar{\mathcal{A}} = \mathrm{fft}(\mathcal{A}, [], 3)$.

Definition 2.1. *(Block diagonal matrix)* $\bar{A}$ *denotes a block diagonal matrix, i.e.,*

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

Definition 2.2. *(Block circulant matrix) For a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its block circulant matrix is defined as following,*

$$\mathtt{bcirc}(\mathcal{A}) = \begin{bmatrix} A^{(1)} & A^{(n_3)} & \cdots & A^{(2)} \\ A^{(2)} & A^{(1)} & \cdots & A^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(n_3)} & A^{(n_3-1)} & \cdots & A^{(1)} \end{bmatrix}.$$

We define the operator

$$\mathtt{unfold}(\mathcal{A}) = \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n_3)} \end{bmatrix}, \mathtt{fold}(\mathtt{unfold}(\mathcal{A})) = \mathcal{A}.$$

Definition 2.3. *(T-product) For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}, \mathcal{B} \in \mathbb{R}^{n_2 \times l \times n_3}$, the t-product $\mathcal{A} * \mathcal{B}$ is defined to be:*

$$\mathcal{A} * \mathcal{B} = \mathtt{fold}(\mathtt{bcirc}(\mathcal{A}) \cdot \mathtt{unfold}(\mathcal{B})). \qquad (5)$$

Note that when $n_3 = 1$ tensor product reduces to matrix product.

Definition 2.4. *(Tensor transpose) The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the $n_2 \times n_1 \times n_3$ tensor $\mathcal{A}^*$ obtained by conjugate transposing each of the frontal slice and then reversing the order of transposed frontal slices 2 through n_3.*

Definition 2.5. *(Identity tensor) The identity tensor $\mathcal{I} \in \mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is the tensor whose frontal slice is the $n \times n$ identity matrix and whose other frontal slices are all zeros.*

Definition 2.6. *(Orthogonal tensor) A tensor $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies*

$$\mathcal{Q}^* * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^* = \mathcal{I}. \qquad (6)$$

Definition 2.7. *(F-diagonal tensor) A tensor is called f-diagonal if each of its frontal slices is a diagonal matrix.*

Thoerem 2.1. *(T-SVD) For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, it can be factored as*

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*, \qquad (7)$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}, \mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is f-diagonal.

Note that

$$(F_{n_3} \bigotimes I_{n_1}) \cdot \mathtt{bcirc}(\mathcal{A}) \cdot (F_{n_3}^{-1} \bigotimes I_{n_2}) = \bar{A}, \qquad (8)$$

where F_{n_3} denotes the $n_3 \times n_3$ Discrete Fourier Transform matrix and $\bigotimes$ denotes Kronecker product. Based on this key property, we can compute tensor SVD efficiently based on matrix SVD in the Fourier domain. One can perform the matrix SVD on each frontal slice of $\bar{\mathcal{A}}$, i.e., $\bar{A}^{(i)} = \bar{U}^{(i)} \bar{S}^{(i)} \bar{V}^{(i)*}$, where $\bar{U}^{(i)}, \bar{S}^{(i)}, \bar{V}^{(i)}$ are frontal slices of $\bar{\mathcal{U}}, \bar{\mathcal{S}}, \bar{\mathcal{V}}$ respectively. Equivalently, $\bar{A} = \bar{U}\bar{S}\bar{V}^*$. After performing `ifft` along the third dimension, we get that $\mathcal{U} = \mathtt{ifft}(\bar{\mathcal{U}}, [], 3), \mathcal{S} = \mathtt{ifft}(\bar{\mathcal{S}}, [], 3), \mathcal{V} = \mathtt{ifft}(\bar{\mathcal{V}}, [], 3)$.

Definition 2.8. *(Tensor multi rank and tubal rank) The multi rank of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a vector $r \in \mathbb{R}^{n_3}$ with its i-th entry as the rank of the i-th frontal slice of $\bar{\mathcal{A}}$, i.e.,*

$r_i = rank(\bar{A}^{(i)})$. The tubal rank $rank_t(\mathcal{A})$ is the number of nonzero singular tubes of $\mathcal{S}$ where $\mathcal{S}$ is from the t-SVD of $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$.

$$rank_t(\mathcal{A}) = \#\{i : \mathcal{S}(i,i,:)\} = \max_i r_i. \tag{9}$$

Definition 2.9. *(Tensor nuclear norm) For a tensor* $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, *its nuclear norm* $||\mathcal{A}||_*$ *is defined as the average of the nuclear norm of all the frontal slices of* $\bar{\mathcal{A}}$, *i.e.,*

$$||\mathcal{A}||_* := \frac{1}{n_3} \sum_{i=1}^{n_3} ||\bar{A}^{(i)}||_*.$$

Note that

$$
\begin{aligned}
||\mathcal{A}||_* &= \frac{1}{n_3} \sum_{i=1}^{n_3} ||\bar{A}^{(i)}||_* = \frac{1}{n_3} ||\bar{A}||_* \\
&= \frac{1}{n_3} ||(F_{n_3} \bigotimes I_{n_1}) \cdot \texttt{bcirc}(\mathcal{A} \cdot (F_{n_3}^{-1} \bigotimes I_{n_2}))||_* \\
&= \frac{1}{n_3} ||\texttt{bcirc}(\mathcal{A})||_*.
\end{aligned}
\tag{10}
$$

So the tensor nuclear norm here is the nuclear norm of a new matricization of a tensor with a factor. Block circulant matricization preserves more spacial relationship.

Definition 2.10. *(Tensor spectral norm) For a tensor* $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, *the tensor spectral norm* $||\mathcal{A}||$ *is defined as* $||\mathcal{A}|| := ||\bar{A}||$.

We define the tensor average rank as $rank_a(\mathcal{A}) = \frac{1}{n_3} \sum_{i=1}^{n_3} rank(\bar{A}^{(i)})$, then the tensor nuclear norm is the convex envelop of the tensor average rank within the unit ball of the tensor spectral norm.

Definition 2.11. *(Tensor incoherence conditions)For* $\mathcal{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *with* $rank_t(\mathcal{L}_0) = r$ *and skinny t-SVD* $\mathcal{L}_0 = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$,*where* $\mathcal{U} \in \mathbb{R}^{n_1 \times r \times n_3}, \mathcal{V} \in \mathbb{R}^{n_2 \times r \times n_3}$ *satisfying* $\mathcal{U}^* * \mathcal{U} = \mathcal{I}, \mathcal{V}^* * \mathcal{V} = \mathcal{I}$ *and* $\mathcal{S} \in \mathbb{R}^{r \times r \times n_3}$*is f-diagonal. Then we say that* $\mathcal{L}_0$ *is said to satisfy the tensor incoherence conditions with parameter* μ *if*

$$\max_{i=1,\cdots,n_1} ||\mathcal{U}^* * \mathring{\mathfrak{e}}_i||_F \le \sqrt{\frac{\mu r}{n_1 n_3}}, \tag{11}$$

$$\max_{i=1,\cdots,n_1} ||\mathcal{V}^* * \mathring{\mathfrak{e}}_j||_F \le \sqrt{\frac{\mu r}{n_2 n_3}}, \tag{12}$$

$$||\mathcal{U} * \mathcal{V}^*||_\infty \le \sqrt{\frac{\mu r}{n_1 n_2 n_3^2}}. \tag{13}$$

3. MAIN RESULT

The following theorem is the main result of this paper. It shows that under certain assumptions, one can exactly recover the low-rank and sparse components from incomplete measurements by solving (2).

To make this problem meaningful, two conditions must be satisfied: (i) $\mathcal{L}_0$ is not sparse and (ii) $\mathcal{S}_0$ is not low-rank. Here we introduce the tensor incoherence condition [4] to describe (i). For (ii), we generate a random model: each element of $\mathcal{S}_0$ is an element of its support set independently with probability ρ. The signs of all entries in support set are i.i.d. Bernoulli variables which take the value ± 1 with probability

1/2. We call this random distribution as an $i.i.d.Bernoulli$-$Rademacher$ model.

Another condition is about the sampling ratio. Intuitively it is impossible to realize perfect recovery without enough measurements. We give a theoretical lower bound for exact recovery in the following theorem. Naturally this bound is closely related to the degrees of freedom in $(\mathcal{L}_0, \mathcal{S}_0)$. Since $\mathcal{L}_0$ is determined by $\bar{L}_0$, we only need to focus on the degrees of freedom of $\bar{L}_0$. We can fully specify $\bar{L}_0$ using $(n_1 + n_2 - r)rn_3$ real numbers. Roughly we can consider $||\mathcal{S}_0||_0$ to be the number of degrees of freedom in $\mathcal{S}_0$.

Thoerem 3.1. *Let* $\mathcal{L}_0, \mathcal{S}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *with* $n_1 = n_2$, *and suppose that* $\mathcal{L}_0$ *is a rank-*r,μ*-incoherent tensor with*

$$r \le \frac{c_r n_2 n_3}{\mu (\log(n_1 n_3)^2}, \tag{14}$$

and $sign(\mathcal{S}_0)$ *is i.i.d. Bernoulli-Rademacher with non-zero probability* $\rho < c_s$. *Let* $Q \subset \mathbb{R}^{n_1 \times n_2 \times n_3}$ *be a random subspace of dimension*

$$dim(Q) \ge C_{meas} \cdot (\rho n_1 n_2 n_3 + n_1 n_3 r) \cdot \log^2(n_1 n_3), \tag{15}$$

distributed according to the Haar measure, probabilistically independent of $sign(\mathcal{S}_0)$. *Then with probability of at least* $1 - C(n_1 n_3)^{-9}$ *in* $(sign(\mathcal{S}_0),Q)$, *the solution to*

$$\min ||\mathcal{L}||_* + \lambda ||\mathcal{S}||_1, \quad s.t. \mathcal{P}_Q[\mathcal{L} + \mathcal{S}] = \mathcal{P}_Q[\mathcal{L}_0 + \mathcal{S}_0], \tag{16}$$

with $\lambda = 1/\sqrt{n_1 n_3}$ *is unique and equal to* $(\mathcal{L}_0, \mathcal{S}_0)$. *Above,* c_r, c_s, c_{meas} *are positive numerical constants.*

[5] gives a tight bound for matrix case. Note that when $n_3 = 1$, our result reduces to [5] so the above theorem is a natural extension of theorem 2.1 in [5].

4. PROOF OF THE MAIN RESULT

In this section, we prove the theorem 3.1. The main strategy is constructing a certificate whose existence leads to the conclusion that $(\mathcal{L}_0, \mathcal{S}_0)$ is the unique optimal solution to the TCPCP problem. In [4], for TPCP problem there exists a certificate Λ_{PCP}. We will use some skills to "upgrade" this certificate to a new certificate for TCPCP problem.

A. Certificate for TCPCP Problem

Here there is no harm in considering two more general problems. Consider a fully observed data tensor $\mathcal{M}$ as a sum of several components:

$$\mathcal{X} = \mathcal{X}_1 + \cdots + \mathcal{X}_\tau, \tag{17}$$

where each $\mathcal{X}_i$ is a low-complexity component. For each type of structure, we use a certain regularizer $|| \cdot ||_{(i)}$ to describe its low complexity. Then we get two more general problems:

$$\min \sum_i \lambda_i ||\mathcal{X}_i||_{(i)}, \quad s.t. \sum_i \mathcal{X}_i = \mathcal{M}, \tag{18}$$

$$\min \sum_i \lambda_i ||\mathcal{X}_i||_{(i)}, \quad s.t. \mathcal{P}_\mathcal{Q}[\sum_i \mathcal{X}_i] = \mathcal{P}_\mathcal{Q}\mathcal{M}. \tag{19}$$

(18) and (19) are general versions of TPCP and TCPCP problems respectively. Here all the $|| \cdot ||_{(i)}$s are required to be *decomposable*.

Definition 4.1. *(Decomposable norm) We say that a norm $||\cdot||$ is locally decomposable at $\mathcal{X}$ if there exists a subspace T and a tensor $\mathcal{S}$ such that*

$$\partial||\cdot||(\mathcal{X}) = \{\Lambda \,|\, \mathcal{P}_T\Lambda = \mathcal{S}, ||\mathcal{P}_{T^\perp}\Lambda||^* \le 1\},$$

where $||\cdot||^*$ denotes the dual norm, and $\mathcal{P}_{T^\perp}$ is non-expansive with respect to $||\cdot||^*$.

For convenience, we introduce two more definitions about certificate:

Definition 4.2. *Λ is an (α,β)-inexact certificate for a putative solution $(\mathcal{X}_1,\cdots,\mathcal{X}_\tau)$ to (18) with parameters $(\lambda_1,\cdots,\lambda_\tau)$ if $||\mathcal{P}_{T_i}\Lambda - \lambda_i\mathcal{S}_i||_F \le \alpha$ and $||\mathcal{P}_{T_i^\perp}\Lambda||^*_{(i)} < \lambda_i\beta$ both hold for each i.*

Definition 4.3. *Λ is an (α,β)-inexact certificate for a putative solution $(\mathcal{X}_1,\cdots,\mathcal{X}_\tau)$ to (19) with parameters $(\lambda_1,\cdots,\lambda_\tau)$ if for each i $||\mathcal{P}_{T_i}\Lambda - \lambda_i\mathcal{S}_i||_F \le \alpha$, $||\mathcal{P}_{T_i^\perp}\Lambda||^*_{(i)} < \lambda_i\beta$, and $\mathcal{P}_{Q^\perp}\Lambda = 0$.*

The following lemma offers a sufficient optimality condition for (19):

Lemma 4.1. *Consider a feasible solution $\mathcal{X} = (\mathcal{X}_1,\cdots,\mathcal{X}_\tau)$ to problem 19. Suppose that each of the norms $||\cdot||_{(i)}$ is locally decomposable at $\mathcal{X}_i$ and that each of the $a_i||\cdot||_{(i)}$ majorizes the Frobenius norm. Then if $T_1,\cdots,T_\tau,\mathcal{Q}^\perp$ are independent subspaces with*

$$||\mathcal{P}_{T_i}\mathcal{P}_{T_j}|| < \frac{1}{\tau - 1} \quad \forall i \ne j, \tag{20}$$

and there exists an (α,β)-inexact certificate $\hat{\Lambda}$, with

$$\beta + \frac{\sqrt{\tau n_3}}{(1 - ||\mathcal{P}_{\mathcal{Q}^\perp}\mathcal{P}_{T_1+\cdots+T_\tau}||^2)\sqrt{1 - (\tau-1)\max_{ij}||\mathcal{P}_{T_i}\mathcal{P}_{T_j}||}}$$
$$\times \frac{\alpha}{\min_l \lambda_l} < 1, \tag{21}$$

the $\mathcal{X}$ is the unique optimal solution.

To prove theorem 3.1, now we only need to construct an inexact certificate with sufficiently small (α,β) as lemma 4.1 shows. The proof of this lemma will be discussed in appendix A.

B. Upgrade

In [4], there exists an inexact certificate for TPCP problem. If $Q = \mathbb{R}^{n_1 \times n_2 \times n_3}$, TCPCP problem reduces to TPCP problem. Intuitively with enough amount of measurements, we may upgrade the certificate for TPCP problem to a new certificate for TCPCP problem with small loss in parameters (α,β).

Thoerem 4.1. *(Upgrade) Consider the general decomposition problem (18), Suppose that each of the norms $||\cdot||_{(i)} \ge ||\cdot||_F/a_i$. Let $\mathcal{X} = (\mathcal{X}_1,\cdots,\mathcal{X}_\tau)$ be feasible for general problem and suppose there exists an (α,β)-inexact certificate $\hat{\Lambda}$ for $\mathcal{X}$ for this problem with parameters (λ_i).*

Then if $Q \subset \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a random subspace distributed according to Haar measure, with

$$dim(Q) \ge C_{sub} \cdot dim(T_1 + \cdots + T_\tau) \cdot \log(n_1 n_3), \tag{22}$$

there exists an inexact certificate for $\mathcal{X}$ with

$$\alpha' \le \alpha + (n_1 n_3)^{-3}||\hat{\Lambda}||_F,$$
$$\beta' \le \beta + C_1 \max_i \frac{\nu_i + a_i\sqrt{\log(n_1 n_3)}}{\lambda_i}\left(\frac{||\hat{\Lambda}||_F^2 \log(n_1 n_3)}{dim(Q)}\right)^{1/2}, \tag{23}$$

with probability at least $1 - C_2 \cdot \tau \cdot (n_1 n_3)^{-9}$ in Q. C_{sub}, C_1, C_2 are all positive numerical constants.

Here the degrees $(n_1 n_3)^{-3}$ and $(n_1 n_3)^{-9}$ can be set to be any larger numbers by appropriate choice of numerical constants. It will be clear in the proof. Compared with upgrade theorem in [3], we introduce factors a_is since tensor nuclear norm can't majorize the Frobenius norm but with a factor. This set is important for offering a tight bound. The proof of this theorem is in appendix B.

C. Certificate for TPCP Problem

The proposition in this section gives an inexact certificate for TPCP problem. This result directly comes from [4] with slight modification. As will become clear in the proof, we can get a certificate for TPCP problem with smaller (α,β) from [4] by appropriate choice of some constants.

Proposition 4.1. *Let $\mathcal{L}_0, \mathcal{S}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Suppose that $\mathcal{L}_0$ is a rank-r tensor satisfying the tensor incoherence conditions with parameter μ where*

$$r \le \frac{\rho_r n_2 n_3}{\mu(\log(n_1 n_3))^2}. \tag{24}$$

Suppose that the support set Ω of $\mathcal{S}_0$ is uniformly distributed with probability ρ where

$$\rho \le \rho_s, \tag{25}$$

and the signs of the non-zero entries are Rademacher random variables. Above, ρ_r, ρ_s are positive numerical constants.

On an event of high probability, the following hold:

$$(i) \quad ||\mathcal{P}_\Omega\mathcal{P}_T|| \le 1/2,$$

and (ii) there exists a $(\frac{1}{(n_1 n_3)^2}, \frac{1}{4})$-inexact TPCP certificate Λ_{PCP} for $(\mathcal{L}_0, \mathcal{S}_0)$, which satisfies

$$||\Lambda_{PCP}||_F \le 4\sqrt{r} + \frac{4}{3}\lambda\sqrt{||\mathcal{S}_0||_0}. \tag{26}$$

The proof is delayed to appendix.

D. Proof of theorem 3.1

Proof. We have $||\cdot||_{(1)} = ||\cdot||_*, ||\cdot||_{(2)} = ||\cdot||_1$. For TCPCP, we take $\lambda_1 = 1, \lambda_2 = 1/\sqrt{n_1 n_3}$. Let $\mathcal{L}_0 = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$ denote the rank-reduced SVD of $\mathcal{L}_0$ and T denote the subspace

$$T = \{\mathcal{U} * \mathcal{X}^* + \mathcal{Y} * \mathcal{V}^* \,|\, \mathcal{X} \in \mathbb{R}^{n_1 \times r \times n_3}, \mathcal{Y} \in \mathbb{R}^{n_2 \times r \times n_3}\}, \tag{27}$$

then the subdifferential of the nuclear norm at $\mathcal{L}_0$ is

$$\partial ||\cdot||_*(\mathcal{L}_0) = \{\Lambda | \mathcal{P}_T\Lambda = \mathcal{U}*\mathcal{V}^*, ||\mathcal{P}_{T^\perp}\Lambda|| \leq 1\}. \quad (28)$$

Let $\Omega = supp(\mathcal{S}_0)$ denote the support of the sparse term. Let $\sum = sign(\mathcal{S}_0)$,then

$$\partial ||\cdot||_1(\mathcal{S}_0) = \{\Lambda | \mathcal{P}_\Omega\Lambda = \sum, ||\mathcal{P}_{\Omega^c}\Lambda||_\infty \leq 1\}. \quad (29)$$

It is easy to check that $||\cdot||_1, ||\cdot||_\infty$ are both decomposable. From lemma 4.1, to show that $(\mathcal{L}_0, \mathcal{S}_0)$ is the unique optimal solution to CPCP problem, it is sufficient to show that

$(I) ||\mathcal{P}_T\mathcal{P}_\Omega|| \leq \dfrac{1}{2}$

(II) There exists an $(\alpha', \dfrac{1}{2})$ − inexact CPCP cerrificate Λ_{CPCP}

$$\text{with} \quad \alpha' \leq (1-||\mathcal{P}_Q\mathcal{P}_{T+\Omega}||^2)/4\sqrt{n_1}n_3. \quad (30)$$

(i) Bounding $1-||\mathcal{P}_Q\mathcal{P}_{T+\Omega}||^2$. We will apply lemma $B.1$, which requires that

$$dim(Q) \geq C_1 \cdot dim(T+\Omega). \quad (31)$$

The dimension of $T+\Omega$ is a random variable, which depends on the size of the support set Ω. Let $\mathcal{E}_\Omega$ denote the event

$$\mathcal{E}_\Omega = \{|\Omega| \leq 2\rho n_1 n_2 n_3 + n_1 n_3\}. \quad (32)$$

Note that $|\Omega|$ is a sum of $n_1 n_2 n_3$ $Ber(\rho)$ random variables. By Bernstein's inequality,

$$\mathbb{P}[|\Omega| \geq \rho n_1 n_2 n_3 + t] \leq \exp(\dfrac{-t^2/2}{\rho n_1 n_2 n_3 + t/3}). \quad (33)$$

Setting $t = \rho n_1 n_2 n_3 + n_1 n_3$, we obtain that

$$\mathbb{P}[\mathcal{E}_\Omega^c] \leq \exp(\dfrac{-3n_1 n_3}{10}). \quad (34)$$

On $\mathcal{E}_\Omega$, we have

$$\begin{aligned} dim(\Omega+T) \\ <2\rho n_1 n_2 n_3 + n_1 n_3 + 2n_1 n_3 r \\ \leq 3\cdot(\rho n_1 n_2 n_3 + n_1 n_3 r). \end{aligned} \quad (35)$$

Comparing with the condition on $dim(Q)$ in the theorem, we can see that on $\mathcal{E}_\Omega$, the condition of Lemma $B.1$ are satisfied. Now, let $S = T+\Omega$, set $B = \{\mathcal{X} \in S\,|\,||\mathcal{X}||_F = 1\}$ and note that

$$\begin{aligned} &1-||\mathcal{P}_{Q^\perp}\mathcal{P}_S||^2 \\ &= \inf_{\mathcal{X}\in B}\langle\mathcal{X},\mathcal{X}\rangle - \langle\mathcal{P}_{Q^\perp}\mathcal{P}_S\mathcal{X}, X\rangle \\ &= \inf_{\mathcal{X}\in B}\langle\mathcal{X}, (\mathcal{P}_S - \mathcal{P}_S\mathcal{P}_{Q^\perp}\mathcal{P}_S)\mathcal{X}\rangle \\ &= \inf_{\mathcal{X}\in B}\langle\mathcal{X}, (\dfrac{dim(Q)}{n_1 n_2 n_3}\mathcal{P}_S + \mathcal{P}_S\mathcal{P}_Q\mathcal{P}_S \\ &\quad -\dfrac{dim(Q)}{n_1 n_2 n_3}\mathcal{P}_S)\mathcal{X}\rangle \\ &\geq \dfrac{dim(Q)}{n_1 n_2 n_3} - \sup_{\mathcal{X}\in B}\langle\mathcal{X}, (\mathcal{P}_S\mathcal{P}_Q\mathcal{P}_S \\ &\quad -\dfrac{dim(Q)}{n_1 n_2 n_3}\mathcal{P}_S)\mathcal{X}\rangle \\ &\geq \dfrac{dim(Q)}{n_1 n_2 n_3} - ||\mathcal{P}_S\mathcal{P}_Q\mathcal{P}_S - \dfrac{dim(Q)}{n_1 n_2 n_3}\mathcal{P}_S||. \end{aligned} \quad (36)$$

Let $\mathcal{E}_Q$ be the event $\{||\mathcal{P}_S\mathcal{P}_Q\mathcal{P}_S - (dim(Q)/n_1 n_2 n_3)\mathcal{P}_S|| \leq \frac{1}{16}(dim(Q)/n_1 n_2 n_3)\}$. Using lemma $B.1$ and $dim(S) \geq n_1 n_3$, we have

$$\mathbb{P}[\mathcal{E}_Q|\mathcal{E}_\Omega] \geq 1 - C_2\exp(-c_1 n_1 n_3). \quad (37)$$

On $\mathcal{E}_Q$,

$$1-||\mathcal{P}_{Q^\perp}\mathcal{P}_S||^2 \geq \dfrac{16}{15}\dfrac{dim(Q)}{n_1 n_2 n_3}. \quad (38)$$

Ensuring that $C_{meas} > \frac{16}{15}$, we can further conclude that

$$1-||\mathcal{P}_{Q^\perp}\mathcal{P}_S||^2 \geq \dfrac{1}{n_1 n_3}. \quad (39)$$

(ii) Inexact TPCP certificate. By theorem 1, on an event $\mathcal{E}_{PCP}$ of probability of at least $1 - C_2(n_1 n_3)^{-10}$, we have $||\mathcal{P}_T\mathcal{P}_\Omega|| < \frac{1}{2}$, and there exists an $((n_1 n_3)^{-2}, \frac{1}{4})$-inexact TPCP certificate Λ_{PCP} for $(\mathcal{L}_0, \mathcal{S}_0)$, with

$$||\Lambda_{PCP}||_F \leq 4\sqrt{r} + 2\lambda\sqrt{|\Omega|}. \quad (40)$$

We rewrite the bound for later use:

$$n_1 n_3 ||\Lambda_{PCP}||_F^2 \leq n_1 n_3(32r + 8\lambda^2|\Omega|), \quad (41)$$

which on $\mathcal{E}_\Omega$ gives

$$n_1 n_3 ||\Lambda_{PCP}||_F^2 \leq C_3(\rho n_1 n_2 n_3 + n_1 n_3 r), \quad (42)$$

where C_3 is numerical.

(iii) Upgrade to TCPCP certificate. Now, condition on $\mathcal{E}_{PCP}$ and $\mathcal{E}_\Omega$. By our assumption on $dim(Q)$ and ensuring C_{meas} is sufficiently large, the conditions of theorem 4.1 are satisfied. On an event $\mathcal{E}_{upgrade}$ of conditional probability of at least $1 - C_4(n_1 n_3)^{-9}$, the certificate Λ_{PCP} can be refined to an (α', β')-inexact TCPCP certificate Λ_{CPCP}, with

$$\alpha' \leq (n_1 n_3)^{-2} + (n_1 n_3)^{-3}||\Lambda_{PCP}||_F \quad (43)$$

and

$$\beta' \leq \dfrac{1}{4} + C_5(\dfrac{||\Lambda_{PCP}||_F^2 \log^2(n_1 n_3)}{dim(Q)})^{1/2}w, \quad (44)$$

where $w = \max\{\mathbb{E}||\mathcal{G}|| + \sqrt{\log(n_1 n_3)}, \mathbb{E}||\mathcal{G}||_\infty\sqrt{n_1 n_3} + \sqrt{\log(n_1 n_3)}\sqrt{n_1 n_3}\}$, $\mathcal{G}$ is i.i.d. $\mathcal{N}(0,1)$. Furthermore, we have the bounds that

$$\mathbb{E}||\mathcal{G}|| \leq 2\sqrt{n_1 n_3} \quad and \quad \mathbb{E}||\mathcal{G}||_\infty \leq 3\sqrt{2\log(n_1 n_3)}. \quad (45)$$

For the first bound, note that $||\mathcal{G}|| = \max_i ||\bar{\mathcal{G}}^{(i)}|| \leq 2||Re(\bar{\mathcal{G}}^{(i)})||$ where $Re(\cdot)$ denotes the real part of matrix. According to the definition of the discrete fourier transform along the 3rd dimension and [9], $||Re(\bar{\mathcal{G}}^{(i)})|| \leq \sqrt{\frac{n_3}{2}}\sqrt{2n_1} = \sqrt{n_1 n_3}$. Thus, we have the first bound. The second can be found on [10]. So,

$$\begin{aligned} \beta' &\leq \dfrac{1}{4} + C_6(\dfrac{||\Lambda_{PCP}||_F^2 n_1 n_3 \log^2(n_1 n_3)}{dim(Q)})^{1/2} \\ &\leq \dfrac{1}{4} + (\dfrac{(\rho n_1 n_2 n_3 + n_1 n_3 r)\log^2(n_1 n_3)}{dim(Q)})^{1/2} \\ &\leq 1/2. \end{aligned} \quad (46)$$

At last, on $\mathcal{E}_\Omega \cap \mathcal{E}_Q$,

$$(1-||\mathcal{P}_{Q^\perp}\mathcal{P}_S||^2)/4\sqrt{n_1}n_3 \geq \dfrac{1}{4n_1^{3/2}n_3^2} \geq \alpha', \quad (47)$$

since we can ensure that $||\Lambda_{PCP}||_F \le \sqrt{n_1 n_3}$ with propriate selection of related constants.

We have shown that on

$$\mathcal{E}_{good} = \mathcal{E}_\Omega \cap \mathcal{E}_Q \cap \mathcal{E}_{PCP} \cap \mathcal{E}_{upgrade}, \quad (48)$$

$(\mathcal{L}_0, \mathcal{S}_0)$ is the unique optimal solution to the TCPCP problem.

(iv) Probability. We have

$$
\begin{aligned}
&\mathbb{P}[\mathcal{E}_{good}^c]\\
&\le \mathbb{P}[(\mathcal{E}_Q \cap \mathcal{E}_\Omega)^c] + \mathbb{P}[(\mathcal{E}_{upgrade} \cap \mathcal{E}_{PCP} \cap \mathcal{E}_\Omega)^c]\\
&= 1 - \mathbb{P}[\mathcal{E}_Q | \mathcal{E}_\Omega]\mathbb{P}[\mathcal{E}_\Omega] + 1\\
&\quad - \mathbb{P}[\mathcal{E}_{upgrade} | \mathcal{E}_{PCP} \cap \mathcal{E}_\Omega]\mathbb{P}[\mathcal{E}_{PCP} \cap \mathcal{E}_\Omega]\\
&\le 1 - \mathbb{P}[\mathcal{E}_Q | \mathcal{E}_\Omega] + \mathbb{P}[\mathcal{E}_\Omega^c] + 1\\
&\quad - \mathbb{P}[\mathcal{E}_{upgrade} | \mathcal{E}_{PCP} \cap \mathcal{E}_\Omega] + \mathbb{P}[\mathcal{E}_{PCP}^c] + \mathbb{P}[\mathcal{E}_\Omega^c]\\
&\le C_2 \exp(-c_1 n_1 n_3) + 2\exp(\frac{-3n_1 n_3}{10}) + C_4 (n_1 n_3)^{-9}\\
&\quad + C_2 (n_1 n_3)^{-10}\\
&\le C(n_1 n_3)^{-9}.
\end{aligned}
\quad (49)
$$

$\square$

5. NUMERICAL EXPERIMENTS

In this section, we design proper algorithm to solve the solution of TCPCP given in theorem 3.1. Then conduct numerical experiments to corroborate our main results, applying TCPCP to image observed by Guassian sampling. Specifically, we solve (16) by Alternating Direction Method of Multipliers (ADMM).

A. Optimization by ADMM

$$\min_{\mathcal{L},\mathcal{S}} \|\mathcal{L}\|_* + \lambda \|\mathcal{S}\|_1, s.t. P_\Omega(\mathcal{S} + \mathcal{L}) = P_\Omega \mathcal{X}_0. \quad (50)$$

Rewrite the constraint of 16 into $G(\mathcal{S}(:) + \mathcal{L}(:)) = G\mathcal{X}_0(:) = g$, where "(:)" denote the MatLab notation. In order to solve this problem by ADMM, we introduce $\mathcal{P}, \mathcal{Q}$ and then obtain the following optimization problem,

$$\min_{\mathcal{S},\mathcal{L},\mathcal{P},\mathcal{Q}} \|\mathcal{P}\|_* + \lambda \|\mathcal{Q}\|_1 + \frac{\mu}{2}\|G(\mathcal{S}(:) + \mathcal{L}(:)) - g\|_F^2,$$
$$s.t. \mathcal{P} = \mathcal{L}, \ \mathcal{Q} = \mathcal{S}.$$

The Lagrangian is augmented as

$$
\begin{aligned}
\mathcal{L}_{\rho_1,\rho_2}(\mathcal{S},\mathcal{L},\mathcal{P},\mathcal{Q}) =& \|P\|_* + \lambda \|\mathcal{Q}\|_1\\
&+ \frac{\rho_1}{2}\|\mathcal{L} - \mathcal{P} + \rho_1^{-1}\mathcal{Z}_1\|_F^2\\
&+ \frac{\rho_2}{2}\|\mathcal{S} - \mathcal{Q} + \rho_2^{-2}\mathcal{Z}_2\|_F^2\\
&+ \frac{\rho_3}{2}\|G(\mathcal{S}(:) + \mathcal{L}(:)) - g + \rho_3^{-2}\mathcal{Z}_3\|_F^2
\end{aligned}
$$

Each of the above sub-problems can be solved as follows,

$$
\begin{aligned}
\mathcal{P} &= arg\min_{\mathcal{P}} \frac{1}{2}\|\mathcal{L} - \mathcal{P} + \rho_1^{-1}\mathcal{Z}_1\|_F^2 + \frac{1}{\rho_1}\|P\|_*\\
&= D_{1/\rho_1}(\mathcal{L} + \rho_1^{-1}\mathcal{Z}_1)\\
\mathcal{Q} &= arg\min_{\mathcal{Q}} \frac{1}{2}\|\mathcal{S} - \mathcal{Q} + \rho_2^{-1}\mathcal{Z}_1\|_F^2 + \frac{\lambda}{\rho_2}\|P\|_*\\
&= S_{\lambda/\rho_2}(\mathcal{S} + \rho_2^{-1}\mathcal{Z}_2)\\
\mathcal{L} &= arg\min_{\mathcal{L}} \frac{\mu}{2}\|G(\mathcal{S}(:) + \mathcal{L}(:)) - g\|_F^2 + \frac{\rho_1}{2}\|\mathcal{L} - \mathcal{P} + \rho_1^{-1}\mathcal{Z}_1\|_F^2\\
&= (\mu G^T G + \rho_1 I)^{-1}(\mu G^T g + \rho_1 \mathcal{P}(:) - \mathcal{Z}_1 - \mu G^T G\mathcal{S}(:))\\
\mathcal{S} &= arg\min_{\mathcal{S}} \frac{\mu}{2}\|G(\mathcal{S}(:) + \mathcal{L}(:)) - g\|_F^2 + \frac{\rho_2}{2}\|\mathcal{S} - \mathcal{Q} + \rho_2^{-1}\mathcal{Z}_2\|_F^2\\
&= (\mu G^T G + \rho_2 I)^{-1}(\mu G^T g + \rho_2 \mathcal{Q}(:) - \mathcal{Z}_2 - \mu G^T G\mathcal{L}(:))
\end{aligned}
$$

Therefore, we can solve the optimization problem (16) by ADMM, shown in Algorithm 1.

Algorithm 1: Solve TCPCP by ADMM

1: Initialization $\mathcal{S} = \mathcal{L} = \mathcal{P} = \mathcal{Q} = \mathcal{Z}_1 = \mathcal{Z}_2 = zeros(dim(X_0))$
2: **for** $k = 1, 2, \cdots$ **do**
3:    update $\mathcal{P}$: $\mathcal{P} = D_{\frac{1}{\rho_1}}(\mathcal{L} + \rho_1^{-1}\mathcal{Z}_1)$
4:    update $\mathcal{Q}$: $\mathcal{P} = S_{\frac{\lambda}{\rho_2}}(\mathcal{S} + \rho_2^{-1}\mathcal{Z}_2)$
5:    update $\mathcal{L}$: $\mathcal{L}(:) = (\mu G^T G + \rho_1 I)^{-1}(\rho_3 G^T g + \rho_1 \mathcal{P}(:) - \mathcal{Z}_1 - \mu G^T G\mathcal{S}(:) - G^T Z_3)$
6:    update $\mathcal{S}$: $\mathcal{S}(:) = (\mu G^T G + \rho_1 I)^{-1}(\rho_3 G^T g + \rho_2 \mathcal{Q}(:) - \mathcal{Z}_2 - \mu G^T G\mathcal{L}(:) - G^T Z_3)$
7:    dual update $\mathcal{Z}_1$: $\mathcal{Z}_1 = \mathcal{Z}_1 + \rho_1 * (\mathcal{L} - \mathcal{P})$
8:    dual update $\mathcal{Z}_2$: $\mathcal{Z}_2 = \mathcal{Z}_2 + \rho_2 * (\mathcal{S} - \mathcal{Q})$
9:    dual update $\mathcal{Z}_3$: $\mathcal{Z}_3 = \mathcal{Z}_3 + \rho_3 * (G(\mathcal{S}(:) + \mathcal{L}(:)) - g)$
10: **end for**

B. Exact Recovery from Varying Fractions of Errors and Observations

6. CONCLUSION

In this paper, we have proved that under certain conditions, one can exactly recover both low-rank and sparse components from a small set of measurements. Meanwhile we established a theoretical bound about the number of measurements required for perfect recovery. In experiment,......

APPENDIX A
PROOF OF LEMMA 4.1

We need two extra lemmas to prove lemma 4.1.

Lemma A.1. *Consider a feasible solution* $\mathcal{X} = (\mathcal{X}_1, \cdots, \mathcal{X}_\tau)$ *to the* (19)*: Suppose that each of the norms* $\|\cdot\|_{(i)}$ *is locally decomposable at* $\mathcal{X}_i$. *If* $T_1, \cdots, T_\tau, \mathcal{Q}^\perp$ *are independent subspaces and there exists* Λ *satisfying* $\mathcal{P}_{T_i}\Lambda = \lambda_i \mathcal{S}_i$ *and* $\|\mathcal{P}_{T_i^\perp}\Lambda\|_{(i)}^* < \lambda_i$ *for each* i, *and* $\mathcal{P}_{\mathcal{Q}^\perp}\Lambda = 0$, *then* $\mathcal{X}$ *is the unique optimal solution to* (19).

Proof. Let f denote the objective function. Consider a feasible perturbation $\delta = (\Delta_1, \cdots, \Delta_\tau)$, so $\mathcal{P}_\mathcal{Q} \sum_i \Delta_i = 0$. Then for

any $\mathcal{W}_1, \cdots, \mathcal{W}_\tau$ such that $\forall i, \mathcal{W}_i \in \partial || \cdot ||_{(i)}(\mathcal{X}_i)$, according to the definition of subgradient we have

$$f(\mathcal{X} + \delta) \geq f(\mathcal{X}) + \sum_i \lambda_i \langle \mathcal{W}_i, \Delta_i \rangle. \tag{51}$$

By duality of norms, for each i there exists $\mathcal{H}_i \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $||\mathcal{H}_i||_{(i)}^* \leq 1$ and

$$\langle \mathcal{H}_i, \mathcal{P}_{T_i^\perp} \Delta_i \rangle = ||\mathcal{P}_{T_i^\perp} \Delta_i||_{(i)}.$$

Set $\mathcal{W}_i = \mathcal{S}_i + \mathcal{P}_{T_i^\perp} \mathcal{H}_i$. Since $\mathcal{P}_{T_i} \mathcal{W}_i = \mathcal{S}_i$ and $||\mathcal{P}_{T_i^\perp} \mathcal{W}_i||_{(i)}^* \leq 1$,

$$\mathcal{W}_i \in \partial || \cdot ||_{(i)}(\mathcal{X}_i).$$

$$\begin{aligned}
\langle \mathcal{W}_i, \Delta_i \rangle &= \langle \mathcal{P}_{T_i} \mathcal{W}_i, \Delta_i \rangle + \langle \mathcal{P}_{T_i^\perp} \mathcal{W}_i, \Delta_i \rangle \\
&= \langle \mathcal{P}_{T_i} \mathcal{W}_i, \mathcal{P}_{T_i} \Delta_i \rangle + \langle \mathcal{P}_{T_i^\perp} \mathcal{W}_i, \mathcal{P}_{T_i^\perp} \Delta_i \rangle \\
&= \langle \mathcal{S}_i, \mathcal{P}_{T_i} \Delta_i \rangle + \langle \mathcal{P}_{T_i^\perp} \mathcal{H}_i, \mathcal{P}_{T_i^\perp} \Delta_i \rangle \tag{52} \\
&= \langle \mathcal{S}_i, \mathcal{P}_{T_i} \Delta_i \rangle + \langle \mathcal{H}_i, \mathcal{P}_{T_i^\perp} \Delta_i \rangle \\
&= \langle \mathcal{S}_i, \mathcal{P}_{T_i} \Delta_i \rangle + ||\mathcal{P}_{T_i^\perp} \Delta_i||_{(i)}.
\end{aligned}$$

$$\begin{aligned}
f(\mathcal{X} + \delta) &\geq f(\mathcal{X}) + \sum_i \langle \lambda_i \mathcal{S}_i, \mathcal{P}_{T_i} \Delta_i \rangle + \lambda_i ||\mathcal{P}_{T_i^\perp} \Delta_i||_i \\
&= f(\mathcal{X}) + \sum_i \langle \mathcal{P}_{T_i} \Lambda, \mathcal{P}_{T_i} \Delta_i \rangle + \lambda_i ||\mathcal{P}_{T_i^\perp} \Delta_i||_i \\
&= f(\mathcal{X}) + \sum_i \langle \Lambda, \mathcal{P}_{T_i} \Delta_i \rangle + \lambda_i ||\mathcal{P}_{T_i^\perp} \Delta_i||_i \\
&= f(\mathcal{X}) + \sum_i \langle \Lambda, \Delta_i \rangle - \langle \mathcal{P}_{T_i^\perp} \Lambda, \mathcal{P}_{T_i^\perp} \Delta_i \rangle \\
&\quad + \lambda_i ||\mathcal{P}_{T_i^\perp} \Delta_i||_i \\
&\geq f(\mathcal{X}) + \sum_i \langle \Lambda, \Delta_i \rangle - ||\mathcal{P}_{T_i^\perp} \Lambda||_i^* ||\mathcal{P}_{T_i^\perp} \Delta_i||_i \\
&\quad + \lambda_i ||\mathcal{P}_{T_i^\perp} \Delta_i||_i \\
&\geq f(\mathcal{X}) + \sum_i (\lambda_i - ||\mathcal{P}_{T_i^\perp} \Lambda||_i^*) ||\mathcal{P}_{T_i^\perp} \Delta_i||_i,
\end{aligned} \tag{53}$$

where we have used that $\Lambda \in \mathcal{Q}, \Delta_j \in \mathcal{Q}^\perp$. Since $||\mathcal{P}_{T_i^\perp} \Lambda||_{(i)}^* < \lambda_i$ for each i, if any of the $||\mathcal{P}_{T_i^\perp} \Delta_i||_i$ is non-zero, then $f(\mathcal{X} + \delta) \geq f(\mathcal{X})$. If $\mathcal{P}_{T_i^\perp} \Delta_i = 0$ for each i, $\Delta_i \in T_i$ which implies that $\sum_i \Delta_i \in (T_1 + \cdots + T_\tau) \cap Q^\perp$. If $\sum_i \Delta_i \neq 0$, this contradicts independence of $(T_1, \cdots, T_\tau, \mathcal{Q}^\perp)$. If $\sum_i \Delta_i = 0$, this contradicts independence of $(T_1, \cdots, T_\tau)$. □

Lemma A.2. *Let $T_1, \cdots, T_k$ be independent subspaces of $\mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{S}_1 \in T_1, \cdots, \mathcal{S}_k \in T_k$. Then the system of equations*

$$\mathcal{P}_{T_i} \mathcal{X} = \mathcal{S}_i, \quad i = 1, \cdots, k \tag{54}$$

has a solution $\mathcal{X} \in T_1 + \cdots + T_k$ satisfying

$$||\mathcal{X}||_F \leq \sqrt{\frac{\sum_i ||\mathcal{S}_i||_F^2}{1 - (k-1) \max_{i \neq j} ||\mathcal{P}_{T_i} \mathcal{P}_{T_j}||}} \tag{55}$$

Proof. Let $vec: \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{n_1 n_2 n_3}$ denote the operator that vectorizes a tensor by stacking its fibers. For each i, let $U_i \in \mathbb{R}^{n_1 n_2 n_3 \times \dim(T_i)}$ denote a matrix whose columns form

an orthonormal basis for $vec[T_i]$. The system of equations is equivalent to

$$\begin{bmatrix} U_1^* \\ \vdots \\ U_k^* \end{bmatrix} x = \begin{bmatrix} U_1^* \cdot vec[S_1] \\ \vdots \\ U_k^* \cdot vec[S_k] \end{bmatrix},$$

where $x = vec[\mathcal{X}]$. Let U^* denote the matrix on the left side and s denote the vector on the right side. The system of equations has a solution $x \in range(U) = vec[T_1 + \cdots + T_k]$ with the l^2 norm at most $||s||_2 / \sigma_{\min}(U)$. Write

$$\begin{pmatrix} I & U_1^* U_2 & \cdots & U_1^* U_k \\ U_2^* U_1 & I & \cdots & U_2^* U_k \\ \vdots & \vdots & \ddots & \vdots \\ U_k^* U_1 & U_k^* U_2 & \cdots & I \end{pmatrix}.$$

Let λ be any eigenvalue of $U^* U$, with corresponding eigenvector $x = (x_1^*, x_2^*, \cdots, x_k^*)^*$. Let $p = \arg\max_j ||x_j||_2$. Then, looking at the p-th block of the equation $\lambda x = UU^* x$, we have

$$\begin{aligned}
|\lambda - 1| \, ||x_p||_2 &= || \sum_{j \neq p} U_p^* U_j x_j ||_2 \\
&\leq \sum_{j \neq p} ||U_p^* U_j|| \, ||x_j||_2 \tag{56} \\
&\leq ||x_p||_2 \times (k-1) \max_{i \neq j} ||U_i^* U_j||.
\end{aligned}$$

Since $||U_i^* U_j|| = ||\mathcal{P}_{T_i} \mathcal{P}_{T_j}||$, we conclude that

$$\sigma_{\min}(U) = \sqrt{\lambda_{\min}(U_* U)} \geq \sqrt{1 - (k-1) \max_{i \neq j} ||\mathcal{P}_{T_i} \mathcal{P}_{T_j}||}. \tag{57}$$

Combined with $||x||_2 \leq ||s||_2 / \sigma_{\min}(U)$, (55) is proved. □

Proof. According to lemma A.2, the system of equations

$$\mathcal{P}_{T_i} \Delta = \lambda_i \mathcal{S}_i - \mathcal{P}_{T_i} \hat{\Lambda}, \quad i = 1, \cdots, \tau \tag{58}$$

has a solution $\Delta_0 \in T_1 + \cdots + T_\tau$ with

$$||\Delta_0||_F \leq \sqrt{\frac{\sum_i ||\lambda_i \mathcal{S}_i - \mathcal{P}_{T_i} \hat{\Lambda}||_F^2}{1 - (\tau-1) \max_{i \neq j} ||\mathcal{P}_{T_i} \mathcal{P}_{T_j}||}}. \tag{59}$$

Since $T_1 + \cdots + T_\tau$ and $\mathcal{Q}^\perp$ are independent, the system of equations

$$\mathcal{P}_{T_1 + \cdots + T_\tau} \Delta = \Delta_0, \quad \mathcal{P}_{\mathcal{Q}^\perp} \Delta = 0 \tag{60}$$

is feasible. We consider a solution Δ_* of minimum Frobenius norm:

$$\Delta_* = \mathcal{P}_{\mathcal{Q}} \sum_{i=1}^\infty (\mathcal{P}_{T_1 + \cdots + T_\tau} \mathcal{P}_{\mathcal{Q}^\perp} \mathcal{P}_{T_1 + \cdots + T_\tau})^i \Delta_0, \tag{61}$$

whose norm is bounded as

$$||\Delta_*||_F \leq \frac{||\Delta_0||_F}{1 - ||\mathcal{P}_{T_1 + \cdots + T_\tau} \mathcal{P}_{\mathcal{Q}^\perp}||^2}. \tag{62}$$

Set $\Lambda = \hat{\Lambda} + \Delta_*$. Note that $\mathcal{P}_{T_i}\Lambda = \lambda_i \mathcal{S}_i$. For each i, we have

$$
\begin{aligned}
\lambda_i^{-1} ||\mathcal{P}_{T_i^\perp}\Lambda||_{(i)}^* &\leq \lambda_i^{-1}||\mathcal{P}_{T_i^\perp}\hat{\Lambda}||_{(i)}^* + \lambda_i^{-1}||\mathcal{P}_{T_i^\perp}\Delta_*||_{(i)}^* \\
&\leq \lambda_i^{-1}||\mathcal{P}_{T_i^\perp}\hat{\Lambda}||_{(i)}^* + \lambda_i^{-1}\sqrt{n_3}||\mathcal{P}_{T_i^\perp}\Delta_*||_F \\
&< \beta + \frac{\sqrt{\tau n_3}}{(1 - ||\mathcal{P}_{\mathcal{Q}^\perp}\mathcal{P}_{T_1+\cdots+T_\tau}||^2)} \\
&\quad \cdot \frac{\alpha}{\sqrt{1 - (\tau-1)\max_{ij}||\mathcal{P}_{T_i}\mathcal{P}_{T_j}||}\min_l \lambda_l} \\
&< 1.
\end{aligned}
\tag{63}
$$

We also note that $\mathcal{P}_{\mathcal{Q}^\perp}\Lambda = 0$. According to lemma A.1, $\mathcal{X}$ is the unique optimal solution. □

APPENDIX B
PROOF OF THEOREM 4.1

Before introducing the proof of upgrade theorem, we prove two important probabilistic lemmas which will be used in later proof.

Lemma B.1. *Let $S \subseteq \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a fixed linear subspace and let $\mathcal{A} = \sum_{j=1}^{\gamma} \mathcal{H}_j\langle\mathcal{H}_j, \cdot\rangle$, where $(\mathcal{H}_j)$ is a sequence of independent i.i.d $\mathcal{N}(0, 1/n_1 n_2 n_3)$ random tensors, and let $\mathcal{R} = range(\mathcal{A}) \subseteq \mathbb{R}^{n_1 \times n_2 \times n_3}$. Then if*

$$
\gamma \geq C_1 \cdot \dim(S),
\tag{64}
$$

with probability of at least $1 - C_2 \exp(-c\gamma)$, for some numerical constants C_1, C_2, c,

$$
\begin{aligned}
&||\mathcal{P}_S \frac{n_1 n_2 n_3}{\gamma}\mathcal{A}\mathcal{P}_S - \mathcal{P}_S|| \leq \frac{1}{2}, \\
&||\mathcal{P}_S\mathcal{P}_R\mathcal{P}_S - \frac{\gamma}{n_1 n_2 n_3}\mathcal{P}_S|| \leq \frac{1}{16}\frac{\gamma}{n_1 n_2 n_3}.
\end{aligned}
\tag{65}
$$

Proof. Fix an $\frac{1}{4}$-net for the unit ball of $||\cdot||_F$, restricted to S. By lemma 5.2 in [6], there exists such a net of size at most $\exp(\dim(S)\log 9)$.
Let $\mathcal{H} : \mathbb{R}^\gamma \to \mathbb{R}^{n_1 \times n_2 \times n_3}$ via $\mathcal{H}\mathbf{x} = \sum_{\mathbf{i}=1}^{\gamma}\mathcal{H}_{\mathbf{i}}\mathbf{x}_{\mathbf{i}}$, and let $\psi : \mathbb{R}^\gamma \to \mathbb{R}^{n_1 \times n_2 \times n_3}$ via $\psi\mathbf{x} = \sum_{\mathbf{i}=1}^{\gamma}\bar{\mathcal{H}}_{\mathbf{i}}\mathbf{x}_{\mathbf{i}}$, where $(\bar{\mathcal{H}}_{\mathbf{i}})$ is an orthonormal sequence of tensors that span R. We have $\mathcal{A} = \mathcal{H}\mathcal{H}^*$ and $\mathcal{P}_R = \psi\psi^*$. According to lemma 5.4 in [6],

$$
\begin{aligned}
&||\mathcal{P}_S \frac{n_1 n_2 n_3}{\gamma}\mathcal{A}\mathcal{P}_S - \mathcal{P}_S|| \\
&= \sup_{\mathcal{X}\in S, ||\mathcal{X}||_F=1} |\langle(\frac{n_1 n_2 n_3}{\gamma}\mathcal{A} - \mathcal{I})\mathcal{X}, \mathcal{X}\rangle| \\
&= \sup_{\mathcal{X}\in S, ||\mathcal{X}||_F=1} |\frac{n_1 n_2 n_3}{\gamma}||\mathcal{H}^*\mathcal{X}||_2^2 - 1| \\
&\leq 2\sup_{\mathcal{X}\in\Gamma} |\frac{n_1 n_2 n_3}{\gamma}||\mathcal{H}^*\mathcal{X}||_2^2 - 1|
\end{aligned}
\tag{66}
$$

After short calculation, we note that $\sqrt{(n_1 n_2 n_3/\gamma)}\mathcal{H}^*\mathcal{X}$ is distributed as an i.i.d $\mathcal{N}(0, 1/\gamma)$ random vector. According to lemma 1 in [7],

$$
\mathbb{P}[|\frac{n_1 n_2 n_3}{\gamma}||\mathcal{H}^*\mathcal{X}||_2^2 - 1| \geq 2\sqrt{\frac{t}{\gamma}} + 2\frac{t}{\gamma}] \leq 2e^{-t}.
\tag{67}
$$

choose $t = c_1\gamma$, with c_1 small enough that $2\sqrt{c_1} + c_1 \leq \frac{1}{4}$. Take a union bound to get

$$
\begin{aligned}
&\mathbb{P}[||\mathcal{P}_S \frac{n_1 n_2 n_3}{\gamma}\mathcal{A}\mathcal{P}_S|| \geq \frac{1}{2}] \\
&\leq \mathbb{P}[\sup_{\mathcal{X}\in\Gamma} |\frac{n_1 n_2 n_3}{\gamma}||\mathcal{H}^*\mathcal{X}||_2^2 - 1| \geq \frac{1}{4}] \\
&\leq 2\exp(-c_1\gamma + \dim(S)\log 9).
\end{aligned}
\tag{68}
$$

Using the assumption $\gamma > C_1\dim(S)$ and ensuring $C_1 > \log 9/c_1$.
For the second term, we use similar strategies to get that

$$
||\frac{n_1 n_2 n_3}{\gamma}\mathcal{P}_S\mathcal{P}_R\mathcal{P}_S - \mathcal{P}_S|| \leq 2\sup_{\mathcal{X}\in\Gamma} |\frac{n_1 n_2 n_3}{\gamma}||\psi^*\mathcal{X}||_2^2 - 1|.
\tag{69}
$$

Note that $\psi^*\mathcal{X}$ is equal in distribution to the restriction of a uniformly distributed random unit vector $r \in \mathbb{S}^{n_1 n_2 n_3 - 1}$ to its first γ coordinates. According to lemma 2.2 of [4], we have that for every $t > 0$, there exits $c_t > 0$ such that

$$
\mathbb{P}[|\frac{n_1 n_2 n_3}{\gamma}||\psi^*\mathcal{X}||_2^2 - 1| > t] \leq \exp(-c_t\gamma).
\tag{70}
$$

Set $t = \frac{1}{32}$. Ensuring C_1 is larger that $\log 9/c_t$ and taking a union bound to get

$$
\begin{aligned}
&\mathbb{P}[\mathcal{P}_S\mathcal{P}_R\mathcal{P}_S - \frac{\gamma}{n_1 n_2 n_3}\mathcal{P}_S|| > \frac{1}{16}\frac{\gamma}{n_1 n_2 n_3}] \\
&\leq \exp(-c_t\gamma + \dim(S)\log 9).
\end{aligned}
\tag{71}
$$

□

Lemma B.2. *Let S be any fixed subspace of $\mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{M}$ be any fixed tensor. Let $\mathcal{A} = \sum_{l=1}^{\gamma} \mathcal{H}_l\langle\mathcal{H}_l, \cdot\rangle$ be a random semidefinite operator constructed from a sequence of independent i.i.d. $\mathcal{N}(0, 1/n_1 n_2 n_3)$ tensors $\mathcal{H}_1, \cdots, \mathcal{H}_\gamma$. Let $||\cdot||$ be any norm that majorizes the $||\cdot||_F/a$ and let $||\cdot||^*$ be its dual norm, where a is a positive numerical constant. Set $\nu = \mathbb{E}[||\mathcal{G}||^*]$, with $\mathcal{G}$ i.i.d. $\mathcal{N}(0, 1)$. Then we have*

$$
||\mathcal{P}_{S^\perp} \frac{n_1 n_2 n_3}{\gamma}\mathcal{A}\mathcal{P}_S\mathcal{M}||^* \leq 10||\mathcal{P}_S\mathcal{M}||_F \frac{\nu + a\sqrt{\log(n_1 n_3)}}{\sqrt{\gamma}},
\tag{72}
$$

with probability of at least $1 - (n_1 n_3)^{-10} - \exp(-\gamma/2)$.

Proof. Let $\Gamma = \{\mathcal{N} \mid ||\mathcal{N}|| \leq 1\} \subset \{\mathcal{N} \mid ||\mathcal{N}||_F \leq a\}$ denote the unit ball for $||\cdot||$. Then

$$
||\mathcal{P}_{S^\perp}\mathcal{A}\mathcal{P}_S\mathcal{M}||^* = \sup_{\mathcal{N}\in\Gamma}\langle\mathcal{N}, \mathcal{P}_{S^\perp}\mathcal{A}\mathcal{P}_S\mathcal{M}\rangle.
\tag{73}
$$

Since S and $S^\perp$ are orthogonal and $\mathcal{H}_l$ is i.i.d. Gaussian, $\mathcal{P}_S\mathcal{H}_l$ and $\mathcal{P}_{S^\perp}\mathcal{H}_l$ are probabilistically independent. So, letting $\mathcal{H}'_1, \cdots, \mathcal{H}'_\gamma$ denote an independent copy of $\mathcal{H}_1, \cdots, \mathcal{H}_\gamma$, we have

$$
\begin{aligned}
\mathcal{P}_{S^\perp}\mathcal{A}\mathcal{P}_S[\cdot] &= \mathcal{P}_{S^\perp}\sum_l \mathcal{H}_l\langle\mathcal{H}_l, \mathcal{P}_S[\cdot]\rangle \\
&= \sum_l \mathcal{P}_{S^\perp}\mathcal{H}_l\langle\mathcal{P}_S\mathcal{H}_l, \cdot\rangle \\
&= \sum_l \mathcal{P}_{S^\perp}\mathcal{H}_l\langle\mathcal{P}_S\mathcal{H}'_l, \cdot\rangle \\
&\triangleq \mathcal{D},
\end{aligned}
\tag{74}
$$

Hence, we have

$$||\mathcal{P}_{S^\perp}\mathcal{A}\mathcal{P}_S\mathcal{M}||^* = ||\mathcal{D}\mathcal{M}||^*. \quad (75)$$

Conditioned on $\mathcal{H}_1^{'}, \cdots, \mathcal{H}_\gamma^{'}$,

$$\xi_\mathcal{N} \triangleq \langle \mathcal{N}, \mathcal{D}\mathcal{M}\rangle = \sum_l \langle \mathcal{P}_S\mathcal{H}_l^{'}, \mathcal{M}\rangle\langle\mathcal{P}_{S^\perp}\mathcal{N}, \mathcal{H}_l\rangle \quad (76)$$

is zero-mean Gaussian. $||\mathcal{D}\mathcal{M}||^* = \sup_{\mathcal{N}\in\Gamma} \xi_\mathcal{N}$.

$$\mathbb{E}[(\xi_\mathcal{N} - \xi_{\mathcal{N}'})^2|\mathcal{H}_1^{'}, \cdots, \mathcal{H}_\gamma^{'}]$$
$$= \frac{\mathcal{P}_{S^\perp}(\mathcal{N}-\mathcal{N}')_F^2}{n_1n_2n_3} \sum_{l=1}^\gamma \langle\mathcal{P}_S\mathcal{H}_l^{'}, \mathcal{M}\rangle^2$$
$$\le \frac{||\mathcal{N}-\mathcal{N}'||_F^2}{n_1n_2n_3} \sum_{l=1}^\gamma \langle\mathcal{P}_S\mathcal{H}_l^{'}, \mathcal{M}\rangle^2 \quad (77)$$
$$= \frac{\Xi^2||\mathcal{N}-\mathcal{N}'||_F^2}{n_1n_2n_3},$$

where we define that

$$\Xi = (\sum_{l=1}^\gamma \langle\mathcal{H}_l^{'}, \mathcal{P}_S\mathcal{M}\rangle^2)^{\frac{1}{2}}. \quad (78)$$

Consider $\zeta_\mathcal{N} = \langle\mathcal{N}, \mathcal{G}\rangle$ where $\mathcal{G}$ is an i.i.d $\mathcal{N}(0, 1/n_1n_2n_3)$ tensor. From the definition of ν,

$$\mathbb{E}[\sup_{\mathcal{N}\in\Gamma} \xi_\mathcal{N}] = \frac{\nu}{\sqrt{n_1n_2n_3}}. \quad (79)$$

Another calculation shows that

$$\mathbb{E}[(\zeta_\mathcal{N} - \zeta_{\mathcal{N}'})^2] = \frac{||\mathcal{N}-\mathcal{N}'||_F^2}{n_1n_2n_3}. \quad (80)$$

By Slepian's inequality, we have

$$\mathbb{E}[\sup_\mathcal{N} \xi_\mathcal{N} |\mathcal{H}_1^{'}, \cdots, \mathcal{H}_\gamma^{'}] \le \Xi \cdot \mathbb{E}[\sup_\mathcal{N} \zeta_\mathcal{N}] = \frac{\nu\Xi}{\sqrt{n_1n_2n_3}}. \quad (81)$$

For fixed $\mathcal{H}_1^{'}, \cdots, \mathcal{H}_\gamma^{'}$ and $\mathcal{N}\in\Gamma$,

$$|\xi_\mathcal{N}(\mathcal{H}_1^{'}, \cdots, \mathcal{H}_\gamma^{'}) - \xi_\mathcal{N}(\tilde{\mathcal{H}}_1^{'}, \cdots, \tilde{\mathcal{H}}_\gamma^{'})|$$
$$= |\sum_{l=1}^\gamma \langle\mathcal{P}_S\mathcal{H}_l^{'}, \mathcal{M}\rangle\langle\mathcal{P}_{S^\perp}\mathcal{N}, \mathcal{H}_l - \tilde{\mathcal{H}}_l\rangle|$$
$$\le (\sum_l \langle\mathcal{P}_S\mathcal{H}_l^{'}, \mathcal{M}\rangle^2)^{\frac{1}{2}}(\sum_l \langle\mathcal{P}_{S^\perp}\mathcal{N}, \mathcal{H}_l - \tilde{\mathcal{H}}_l\rangle^2)^{\frac{1}{2}} \quad (82)$$
$$\le a\Xi(\sum_l ||\mathcal{H}_l - \tilde{\mathcal{H}}_l^{'}||_F^2)^{\frac{1}{2}}.$$

$\xi_\mathcal{N}$ is a $a\Xi-$ Lipschitz function of the i.i.d. Gaussian sequence $(\mathcal{H}_1, \cdots, \mathcal{H}_\gamma)$. Hence, the supremum $||\mathcal{D}\mathcal{M}^*||$ is also $a\Xi-$ Lipschitz. By proposition 2.18 in [8],

$$\mathbb{P}[||\mathcal{D}\mathcal{M}||^* > \mathbb{E}[||\mathcal{D}\mathcal{M}||^* | (\mathcal{H}_l^{'})] + \frac{ta\Xi}{\sqrt{n_1n_2n_3}}] \le \exp(-\frac{t^2}{2}). \quad (83)$$

Combining with previous estimates, we have

$$\mathbb{P}[||\mathcal{D}\mathcal{M}||^* > \frac{\Xi(\mu + at)}{\sqrt{n_1n_2n_3}}|(\mathcal{H}_l^{'})] \le \exp(-\frac{t^2}{2}). \quad (84)$$

Since this estimate holds for any $\mathcal{H}_l^{'}$, it holds unconditionally:

$$\mathbb{P}[||\mathcal{D}\mathcal{M}||^* > \frac{\Xi(\mu + at)}{\sqrt{n_1n_2n_3}}] \le \exp(-\frac{t^2}{2}). \quad (85)$$

Note that Ξ is a $||\mathcal{P}_S\mathcal{M}||_F$-Lipschitz function of i.i.d. Gaussian sequence $(\mathcal{H}_1^{'}, \cdots, \mathcal{H}_\gamma^{'})$, with

$$|\Xi(\mathcal{H}_1^{'}, \cdots, \mathcal{H}_\gamma^{'}) - \Xi(\tilde{\mathcal{H}}_1^{'}, \cdots, \tilde{\mathcal{H}}_\gamma^{'})|$$
$$= |\sqrt{\sum_{l=1}^r \langle\mathcal{P}_S\mathcal{H}_l^{'}, \mathcal{M}\rangle^2} - \sqrt{\sum_{l=1}^r \langle\mathcal{P}_S\tilde{\mathcal{H}}_l^{'}, \mathcal{M}\rangle^2}|$$
$$\le \sqrt{\sum_{l=1}^r \rangle\mathcal{P}_S(\tilde{\mathcal{H}}_l^{'} - \mathcal{H}_l^{'}), \mathcal{M}\rangle^2} \quad (86)$$
$$\le ||\mathcal{P}_S\mathcal{M}||_F \sqrt{\sum_{l=1}^r ||\mathcal{H}_l^{'} - \tilde{\mathcal{H}}_l^{'}||^2}.$$

Moreover,

$$\mathbb{E}(\Xi) \le (\mathbb{E}[\Xi^2])^{1/2} = ||\mathcal{P}_S\mathcal{M}||_F \frac{\gamma}{n_1n_2n_3}. \quad (87)$$

From Lipschitz concentration,

$$\mathbb{P}[\Xi > \mathbb{E}[\Xi] + s||\mathcal{P}_S\mathcal{M}||_F] \le \exp(-\frac{s^2n_1n_2n_3}{2}), \quad (88)$$

let $s = \sqrt{\frac{\gamma}{n_1n_2n_3}}$ we can get that

$$\mathbb{P}[\Xi > 2||\mathcal{P}_S\mathcal{M}||_F \sqrt{\frac{\gamma}{n_1n_2n_3}}] \le \exp(-\frac{\gamma}{2}). \quad (89)$$

Above all,

$$\mathbb{P}[||\frac{n_1n_2n_3}{\gamma}\mathcal{P}_{S^\perp}\mathcal{A}\mathcal{P}_S\mathcal{M}||^*$$
$$> 10||\mathcal{P}_S\mathcal{M}||_F \frac{\nu + \sqrt{\log(n_1n_3)}a}{\sqrt{\gamma}}] \quad (90)$$
$$\le (n_1n_3)^{-10} + \exp(-\frac{\gamma}{2}).$$

$\square$

Proof. Consider the space $S = T_1 + \cdots + T_\tau + span(\tilde{\Lambda})$, then S is a linear subspace of dimension at most $dim(T_1 + \cdots + T_\tau) + 1$ containing $\tilde{\Lambda}$. We will generate a new certificate Λ_*. The new certificate should inherit good properties of $\hat{\Lambda}$ on $T_1 + \cdots + T_\tau$ and satisfies

$$\mathcal{P}_{Q^\perp}\Lambda_* = 0. \quad (91)$$

To this end, we will set

$$\Lambda_0 = 0. \quad (92)$$

We will generate inductively a sequence $(\Lambda_j)_{j=1, \cdots, k}$ for appropriate k, such that with high probability $\Lambda_* = \Lambda_k$ is the desired certificate. Define the error at step j to be

$$\mathcal{E}_j = \mathcal{P}_S[\Lambda_j] - \hat{\Lambda} \in S. \quad (93)$$

We will generate a sequence of corrections which drive $\mathcal{E}_j$ towards zero.

By orthogonal invariance, Q is equal in distribution to the linear span of $\mathcal{H}_1, \cdots, \mathcal{H}_{dim(Q)}$, where $\mathcal{H}_j$ are independent i.i.d. $\mathcal{N}(0, 1/n_1n_2n_3)$ random tensors. Choose from $1, \cdots, dim(Q)$,

$$k = \lceil 3\log_2(n_1n_3)\rceil, \quad (94)$$

disjoint subsets $I_1, \cdots, I_k$ of size

$$\gamma = \lfloor \frac{dim(Q)}{k} \rfloor. \tag{95}$$

Our choice of constant ensures that $2^{-k} \leq (n_1 n_3)^{-3}$. we can make sure that

$$\gamma \geq C_3 \cdot dim(S) \tag{96}$$

where C_3 is a numerical constant. Since $dim(Q) \geq C_{sub} \cdot dim(T_1 + \cdots + T_\tau) \cdot \log(n_1 n_3)$, once C_3 is chosen, we ensure that C_{sub} is large enough that $\gamma \geq C_3 \cdot dim(S)$.

Let $\mathcal{A}_j : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{n_1 \times n_2 \times n_3}$ denote the semidefinite operator that acts via

$$\mathcal{A}_j[\cdot] = \sum_{i \in I_j} \mathcal{H}_i \langle \mathcal{H}_i, \cdot \rangle. \tag{97}$$

Note that $\mathbb{E}[\mathcal{A}_j] = (\gamma/n_1 n_2 n_3)\mathcal{I}$. For $j = 1, \cdots, k$, let

$$\begin{aligned}
\Lambda_j &= \Lambda_{j-1} - \frac{n_1 n_2 n_3}{\gamma} \mathcal{A}_j E_{j-1} \\
&= -\sum_{i=1}^{j} \frac{n_1 n_2 n_3}{\gamma} \mathcal{A}_i E_{i-1}.
\end{aligned} \tag{98}$$

Then we have

$$\begin{aligned}
E_j &= \mathcal{P}_S[\Lambda_j] - \hat{\Lambda} \\
&= \mathcal{P}_S[\Lambda_{j-1}] - \hat{\Lambda} - \mathcal{P}_S \frac{n_1 n_2 n_3}{\gamma} \mathcal{A}_j E_{j-1} \\
&= \mathcal{P}_S(\mathcal{I} - \frac{n_1 n_2 n_3}{\gamma} \mathcal{A}_j)\mathcal{P}_S E_{j-1}.
\end{aligned} \tag{99}$$

We may further write

$$\begin{aligned}
\Lambda_k &= \mathcal{P}_S[\Lambda_k] + \mathcal{P}_{S^\perp}[\Lambda_k] \\
&= \hat{\Lambda} + E_k - \sum_{j=1}^{k} \mathcal{P}_{S^\perp} \frac{n_1 n_2 n_3}{\gamma} \mathcal{A}_j \mathcal{P}_S E_{j-1}.
\end{aligned} \tag{100}$$

(i) Driving E to zero. Ensuring that C_3 is sufficiently large that the hypotheses of lemma $B.1$ are verified, we have that with probability of at least $1 - C_4 \exp(-c_1 \gamma)$,

$$\|\mathcal{P}_S(\mathcal{I} - \frac{n_1 n_2 n_3}{\gamma} \mathcal{A}_j)\mathcal{P}_S\| = \|\mathcal{P}_S \frac{n_1 n_2 n_3}{\gamma} \mathcal{A}_j \mathcal{P}_S - \mathcal{P}_S\| \leq \frac{1}{2}. \tag{101}$$

Hence, we have that $\|E_j\|_F \leq \frac{1}{2}\|E_{j-1}\|_F$ for each j, on the complement of a bad event $\mathcal{E}_{err}$ of probability of at most $C_4 k \exp(-c_1 \gamma)$. On $\mathcal{E}_{err}^c$, we have

$$\|E_k\|_F \leq 2^{-k}\|\hat{\Lambda}\|_F, \quad \sum_{j=1}^{k} \|E_j\|_F \leq 2\|\hat{\Lambda}\|_F. \tag{102}$$

(ii)Analysis of α'. From the definition, we may set $\alpha' = \max_i \|\mathcal{P}_{T_i}\Lambda_k - \lambda_i S_i\|_F$. On $\mathcal{E}_{err}^c$,

$$\begin{aligned}
\|\mathcal{P}_{T_i}\Lambda_k - \lambda_i S_i\|_F &= \|\mathcal{P}_{T_i}[\hat{\Lambda} + E_k] - \lambda_i S_i\|_F \\
&\leq \|\mathcal{P}_{T_i}\hat{\Lambda} - \lambda_i S_i\|_F + \|E_k\|_F \\
&\leq \|\mathcal{P}_{T_i}\hat{\Lambda} - \lambda_i S_i\|_F + 2^{-k}\|\hat{\Lambda}\|_F \\
&\leq \|\mathcal{P}_{T_i}\hat{\Lambda} - \lambda_i S_i\|_F + (n_1 n_3)^{-3}\|\hat{\Lambda}\|_F \\
&\leq \alpha + (n_1 n_3)^{-3}\|\hat{\Lambda}\|_F.
\end{aligned} \tag{103}$$

(iii)Analysis of β'.

$$\beta' = \max_{i=1,\cdots,\tau} \lambda_i^{-1} \|\Lambda_k\|_{(i)}^*. \tag{104}$$

From triangle inequality, we have that

$$\begin{aligned}
\|\Lambda_k\|_{(i)}^* &\leq \|\hat{\Lambda}\|_{(i)}^* + \sqrt{n_3}\|E_k\|_F \\
&+ \sum_{j=1}^{k} \|\mathcal{P}_{S^\perp} \frac{n_1 n_2 n_3}{\gamma} \mathcal{A}_j \mathcal{P}_S E_{j-1}\|_{(i)}^*.
\end{aligned} \tag{105}$$

Applying lemma $B.2$, this is bounded by

$$\begin{aligned}
&\|\Lambda_k\|_{(i)}^* \\
&\leq \lambda_i \beta + 2^{-k}\sqrt{n_3}\|E_0\|_F \\
&+ 10\frac{\nu_i + a_i\sqrt{\log(n_1 n_3)}}{\sqrt{\gamma}} \sum_{j=1}^{k} \|E_{j-1}\|_F \\
&\leq \lambda_i \beta + 21\frac{\nu_i + a_i\sqrt{\log(n_1 n_3)}}{\sqrt{\gamma}}\|\hat{\Lambda}\|_F,
\end{aligned} \tag{106}$$

on the complement of an event $\mathcal{E}_\infty$ of probability of at most $k(n_1 n_3)^{-10} + k \exp(-\gamma/2) + \mathbb{P}[\mathcal{E}_{err}]$. Since $\gamma \geq c \cdot dim(Q)/\log(n_1 n_3)$, for some numerical constant C_5,

$$\begin{aligned}
&\lambda_i^{-1}\|\Lambda_k\|_{(i)}^* \\
&\leq \beta + C_5 \frac{\nu_i + a_i\sqrt{\log(n_1 n_3)}}{\lambda_i}(\frac{\log(n_1 n_3)\|\hat{\Lambda}\|_F^2}{dim(Q)})^{1/2}.
\end{aligned} \tag{107}$$

$\square$

APPENDIX C
PROOF OF PROPOSITION 4.1

Under the hypotheses, (i) follows immediately from Corollary 4.2 of [4]. [4] constructs a certificate Λ_{PCP} in three parts as

$$\Lambda_{PCP} = \mathcal{U} * \mathcal{V}^* + \mathcal{W}^{\mathcal{L}} + \mathcal{W}^{\mathcal{S}}.$$

Since $\mathcal{W}^{\mathcal{L}}, \mathcal{W}^{\mathcal{S}} \in T^\perp$,

$$\|\mathcal{P}_T \Lambda_{PCP} - \mathcal{U} * \mathcal{V}^*\|_F = 0.$$

Moreover, $\mathcal{P}_\Omega \mathcal{W}^{\mathcal{S}} = \lambda \, sign(\mathcal{S}_0)$. So, we have

$$\|\mathcal{P}_\Omega \Lambda_{PCP} - \lambda \, sign(\mathcal{S}_0)\|_F = \|\mathcal{P}_\Omega[\mathcal{U} * \mathcal{V}^* + \mathcal{W}^{\mathcal{L}}]\|_F.$$

To prove proposition 4.1, it is suffice to show that with high probability the following properties are satisfied:

(I) Structure constraint:

$$\|\mathcal{P}_\Omega[\mathcal{U} * \mathcal{V}^* + \mathcal{W}^{\mathcal{L}}]\|_F \leq \frac{1}{(n_1 n_3)^2}. \tag{108}$$

(II) Dual norm constraints:

$$\|\mathcal{P}_{T^\perp}\mathcal{W}^{\mathcal{L}}\| \leq \frac{1}{8}, \quad \|\mathcal{P}_{\Omega^c}[\mathcal{U} * \mathcal{V}^* + \mathcal{W}^{\mathcal{L}}]\|_\infty \leq \frac{\lambda}{8} \tag{109}$$

and

$$\|\mathcal{P}_{T^\perp}\mathcal{W}^{\mathcal{S}}\| \leq \frac{1}{8}, \quad \|\mathcal{P}_{\Omega^c}\mathcal{W}^{\mathcal{S}}\|_\infty \leq \frac{\lambda}{8}. \tag{110}$$

(III) Frobenius norm bounds: We have $\|\Lambda_{PCP}\|_F \leq \|\mathcal{U} * \mathcal{V}^*\|_F + \|\mathcal{W}^{\mathcal{L}}\|_F + \|\mathcal{W}^{\mathcal{S}}\|_F$. The first term is $\sqrt{r}$. We will show that

$$\|\mathcal{W}^{\mathcal{L}}\|_F \leq 3\sqrt{r}, \tag{111}$$

$$||\mathcal{W}^{\mathcal{S}}||_F \leq \frac{4}{3}\lambda\sqrt{||\mathcal{S}_0||_0}. \tag{112}$$

In paragraph (i)(ii), we show that with slight changes in the constants, (I)(II) are satisfied with a high probability. In the end of paragraph (ii) and paragraph (iii), we show (111) and (112).

(i) Review of lemma 3.4 in [4].
Set $\mathcal{Z}_j = \mathcal{U} * \mathcal{V}^* - \mathcal{P}_T\mathcal{Y}_j$ obeying

$$\mathcal{Z}_j = (\mathcal{P}_T - q^{-1}\mathcal{P}_T\mathcal{P}_{\Omega_j}\mathcal{P}_T)\mathcal{Z}_{j-1}.$$

Note that when $q \leq C_0\epsilon^{-2}\frac{\mu r \log(n_1 n_3)}{n_2 n_3}$, we have

$$\begin{aligned}||\mathcal{Z}_j||_\infty &\leq \epsilon^j||\mathcal{U} * \mathcal{V}^*||_\infty, \\ ||\mathcal{Z}_j||_F &\leq \epsilon^j||\mathcal{U} * \mathcal{V}^*||_F \leq \epsilon^j\sqrt{r},\end{aligned} \tag{113}$$

by lemma 4.1 and lemma 4.3 in [4]. We assume $\epsilon \leq e^{-1}$. We have

$$||\mathcal{W}^{\mathcal{L}}|| \leq C_1\epsilon, \tag{114}$$

for some numerical constant C_1.
Since $\mathcal{P}_\Omega(\mathcal{U} * \mathcal{V}^* + \mathcal{P}_{T^\perp}\mathcal{Y}_{j_0}) = \mathcal{P}_\Omega(\mathcal{Z}_{j_0})$ and $||\mathcal{Z}_{j_0}||_F \leq \epsilon^{j_0}\sqrt{r}$,

$$\begin{aligned}||\mathcal{P}_\Omega(\mathcal{U} * \mathcal{V}^* + \mathcal{P}_{T^\perp}\mathcal{Y}_{j_0})|| &\leq \epsilon^{j_0}\sqrt{r} \\ &\leq \epsilon^{j_0}\sqrt{\frac{\rho_r n_2 n_3}{\mu(\log(n_1 n_3))^2}}\end{aligned} \tag{115}$$

Now we set $j_0 = 3\lceil\log(n_1 n_3)\rceil$, (108) is achieved.
We have $\mathcal{U} * \mathcal{V}^* + \mathcal{W}^{\mathcal{L}} = \mathcal{Z}_{j_0} + \mathcal{Y}_{j_0}$.
Since $||\mathcal{Z}_{j_0}||_\infty \leq \epsilon^{j_0}||\mathcal{U} * \mathcal{V}^*||_\infty$ and $||\mathcal{U} * \mathcal{V}^*||_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2 n_3^2}}$, combined with (24) we have

$$||\mathcal{Z}_j||_\infty \leq \epsilon^{j_0}\sqrt{\frac{\rho_r}{(\log(n_1 n_3))^2}}\lambda \leq \frac{\lambda}{16}.$$

[4] has established that

$$||\mathcal{Y}_{j_0}||_\infty \leq C_2\frac{\epsilon^2}{\sqrt{\mu r(\log(n_1 n_3))^2}},$$

for some numerical constant C_2.
$||\mathcal{Y}_{j_0}|| \leq \frac{\lambda}{16}$ if

$$\epsilon \leq C_3\left(\frac{\mu r(\log(n_1 n_3))^2}{n_1 n_3}\right)^{1/4}. \tag{116}$$

(109) is proved with sufficiently small ϵ (provided that ρ_r is sufficiently small).

(ii) Review of lemma 3.5 in [4].
Denote $\mathcal{M} = sign(\mathcal{S}_0)$.
By construction,

$$\begin{aligned}\mathcal{W}^{\mathcal{S}} &= \lambda\mathcal{P}_{T^\perp}\mathcal{M} + \lambda\mathcal{P}_{T^\perp}\sum_{k \geq 1}(\mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega)^k\mathcal{M} \\ &:= \mathcal{P}_{T^\perp}\mathcal{W}_0^{\mathcal{S}} + \mathcal{P}_{T^\perp}\mathcal{W}_1^{\mathcal{S}}.\end{aligned} \tag{117}$$

[4] have proved that $||\mathcal{P}_{T^\perp}\mathcal{W}_0^{\mathcal{S}}||, ||\mathcal{P}_{T^\perp}\mathcal{W}_1^{\mathcal{S}}||$ are small enough when ρ is sufficiently small. Therefore $||\mathcal{W}^{\mathcal{S}}|| \leq 1/8$.

[4] also proved that $||\mathcal{P}_{\Omega^\perp}\mathcal{W}^{\mathcal{S}}||_\infty \leq t\lambda$ holds with high probability when ρ_r is sufficiently small. So we set $t = \frac{1}{8}$ here.

$\mathcal{W}^{\mathcal{S}}$ is constructed by Neumann Series

$$\mathcal{W}^{\mathcal{S}} = \lambda\mathcal{P}_{T^\perp}\sum_{k \geq 0}(\mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega)^k\mathcal{M}.$$

So,

$$||\mathcal{W}^{\mathcal{S}}||_F \leq \frac{||\lambda sign(\mathcal{S}_0)||_F}{1 - ||\mathcal{P}_\Omega\mathcal{P}_T||^2} \leq \frac{4}{3}\lambda\sqrt{||\mathcal{S}_0||_0}. \tag{118}$$

(iii) Bounding $||\mathcal{W}^{\mathcal{L}}||_F$. Since $\mathcal{W}^{\mathcal{L}} = \mathcal{P}_{T^\perp}\sum_j q^{-1}\mathcal{P}_{\Omega_j}\mathcal{Z}_{j-1}$, it is enough to control the Frobenius norm $q^{-1}\mathcal{P}_{\Omega_j}\mathcal{Z}_{j-1}$ for each j. Note that

$$\begin{aligned}&||q^{-1}\mathcal{P}_{\Omega_j}\mathcal{Z}_{j-1}||_F^2 \\ =&||\mathcal{Z}_{j-1}||_F^2 + \sum_{kls}(q^{-1}\delta_{kls} - 1)[\mathcal{Z}_{j-1}]_{kls}^2 \\ \doteq&||\mathcal{Z}_{j-1}||_F^2 + \sum_{kls}\mathcal{H}_{kls},\end{aligned} \tag{119}$$

where δ_{kls} is an indicator for the event $(k, l, s) \in \Omega_j$. Then $\mathbb{E}[\mathcal{H}_{kls}] = 0, |\mathcal{H}_{kls}| \leq q^{-1}||\mathcal{Z}_{j-1}||_\infty^2$ almost surely and $\mathbb{E}[\mathcal{H}_{kls}^2] \leq q^{-1}[\mathcal{Z}_{j-1}]_{kls}^4$. Summing, we have

$$\sum_{kls}\mathbb{E}[\mathcal{H}_{kls}^2] \leq q^{-1}||\mathcal{Z}_{j-1}||_\infty^2||\mathcal{Z}_{j-1}||_F^2. \tag{120}$$

By Bernstein's inequality,

$$\begin{aligned}&\mathbb{P}[||q^{-1}\mathcal{P}_{\Omega_j}\mathcal{Z}_{j-1}||_F^2 > ||\mathcal{Z}_{j-1}||_F^2 + t] \\ \leq&\exp(-\frac{t^2}{2(q^{-1}||\mathcal{Z}_{j-1}||_\infty^2||\mathcal{Z}_{j-1}||_F^2 + \frac{1}{3}q^{-1}||\mathcal{Z}_{j-1}||_\infty^2 t)}) \\ =&\exp(-\frac{t^2}{2q^{-1}||\mathcal{Z}_{j-1}||_\infty^2(||\mathcal{Z}_{j-1}||_F^2 + \frac{t}{3})}).\end{aligned} \tag{121}$$

By setting

$$\begin{aligned}t_j = C_4 \max(&||\mathcal{Z}_{j-1}||_\infty^2 q^{-1}\log(n_1 n_3), \\ &||\mathcal{Z}_{j-1}||_\infty||\mathcal{Z}_{j-1}||_F\sqrt{q^{-1}\log(n_1 n_3)}),\end{aligned} \tag{122}$$

with appropriate numerical constant C_4, we can ensure that for each j,

$$[||q^{-1}\mathcal{P}_{\Omega_j}\mathcal{Z}_{j-1}||_F^2 > ||\mathcal{Z}_{j-1}||_F^2 + t_j] \leq (n_1 n_3)^{-11}.$$

Since we have $q > c/\log(n_1 n_3)$ for some positive numerical constant c, on an event with an overall probability of at least

$$1 - j_0(n_1 n_3)^{-11},$$

$$||\mathcal{W}^{\mathcal{L}}||_F$$

$$\leq \sum_{j=1}^{j_0} ||q^{-1}\mathcal{P}_{\Omega_j}\mathcal{Z}_{j-1}||_F$$

$$\leq \sum_{j=1}^{j_0} ||\mathcal{Z}_{j-1}||_F + \sqrt{t_j}$$

$$\leq \sum_{j=1}^{j_0} ||\mathcal{Z}_{j-1}||_F + C_5||\mathcal{Z}_{j-1}||_\infty \log(n_1 n_3)$$

$$+ C_6 \sqrt{||\mathcal{Z}_{j-1}||_\infty ||\mathcal{Z}_{j-1}||_F \log(n_1 n_3)} \qquad (123)$$

$$\leq 2\sqrt{r} + 2C_5 \log(n_1 n_3)\sqrt{\frac{\mu r}{n_1 n_2 n_3^2}}$$

$$+ C_6 \sqrt{\log(n_1 n_3)} \sum_{j=1}^{j_0} \epsilon^j ||\mathcal{Z}_0||_\infty^{1/2} ||\mathcal{Z}_0||_F^{1/2}$$

$$\leq 2\sqrt{r} + 2C_5 \log(n_1 n_3)\sqrt{\frac{\mu r}{n_1 n_2 n_3^2}}$$

$$+ 2C_6 \sqrt{\log(n_1 n_3)} \sqrt[4]{\frac{\mu r^2}{n_1 n_2 n_3^2}}.$$

Recalling (24) and ensuring ρ_r is sufficiently small, we can conclude that $||\mathcal{W}^{\mathcal{L}}||_F \leq 3\sqrt{r}$.

ACKNOWLEDGMENT

REFERENCES

[1] Kolda T G, Bader B W. Tensor decompositions and applications[J]. SIAM review, 2009, 51(3): 455-500.

[2] Kilmer M E, Braman K, Hao N, et al. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging[J]. SIAM Journal on Matrix Analysis and Applications, 2013, 34(1): 148-172.

[3] Cands E J, Li X, Ma Y, et al. Robust principal component analysis?[J]. Journal of the ACM (JACM), 2011, 58(3): 11.

[4] Lu C, Feng J, Chen Y, et al. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 5249-5257.

[5] Wright J, Ganesh A, Min K, et al. Compressive principal component pursuit[J]. Information and Inference: A Journal of the IMA, 2013, 2(1): 32-68.

[6] Vershynin, R. (2011) Introduction to the non-asymptotic analysis of random matrices. Available at http://www-personal.umich.edu/romanv/papers/non-asymptotic-rmt-plain.pdf.

[7] Laurent, B. & Massart, P. (2000) Adaptive estimation of a quadratic function by model selection. Ann. Statist., 28, 1302−1338.

[8] Ledoux, M. (2001) The Concentration of Measure Phenomenon. Providence, RI: American Mathematical Society.

[9] Davidson, K. & Szarek, S. (2001) Local operator theory, random matrices and banach spaces. Handbook of the Geometry of Banach Spaces (W. B. Johnson & J. Lindenstrauss eds) Elsevier Science, pp. 317−366.

[10] Vershynin, R. (2009) Lectures on geometric functional analysis. Available at http://www-personal.umich. edu/ romanv/papers/GFA-book/GFA-book.pdf.

[11] Trmoulhac B, Dikaios N, Atkinson D, et al. Dynamic MR Image ReconstructionSeparation From Undersampled $(\mathbf{k}, t)$-Space via Low-Rank Plus Sparse Prior[J]. IEEE transactions on medical imaging, 2014, 33(8): 1689-1701.

PLACE PHOTO HERE

Michael Shell Biography text here.

John Doe Biography text here.