Supplementary Material of Tensor Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Tensors via Convex Optimization

Canyi Lu[1], Jiashi Feng[1], Yudong Chen[2], Wei Liu[3], Zhouchen Lin[4,5,*], Shuicheng Yan[6,1]
[1] Department of Electrical and Computer Engineering, National University of Singapore
[2] School of Operations Research and Information Engineering, Cornell University [3] Didi Research
[4] Key Laboratory of Machine Perception (MOE), School of EECS, Peking University
[5] Cooperative Medianet Innovation Center, Shanghai Jiaotong University [6] 360 AI Institute
canyilu@gmail.com, elefjia@nus.edu.sg, yudong.chen@cornell.edu, wliu@ee.columbia.edu
zlin@pku.edu.cn, eleyans@nus.edu.sg

July 12, 2016

1 Structure of This Document

This document gives the detailed proof of Theorem 3.1 in the manuscript. Section 2 gives some other notations and properties which will be used in the proofs. Section 3 provides a way for the construction of the solution to the TRPCA problem, and Section 4 proves that the constructed solution is exact to the TRPCA problem. Section 5 gives the proofs of some lemmas which are used in Section 4.

Note that several parts of our proofs are quite different from RPCA [1]. For example, the proofs of Lemma 4.5 and many lemmas in Section 3. Our proofs further consider the properties of Fourier transformation and block circulant matrix.

2 Notations

Beyond the notations introduced in the paper, we need some other notations used in this document. The tensor spectral (or operator) norm of $\boldsymbol{\mathcal{A}}$ is defined as $\|\boldsymbol{\mathcal{A}}\| = \|\bar{\boldsymbol{A}}\|$. The operator norm of an operator on tensor is defined as $\|\mathfrak{L}\| = \sup_{\|\boldsymbol{\mathcal{A}}\|_F=1}\|\mathfrak{L}(\boldsymbol{\mathcal{A}})\|_F$. The inner product of tensors has the following property

$$\langle\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}\rangle = \frac{1}{n_3}\langle\bar{\boldsymbol{A}},\bar{\boldsymbol{B}}\rangle. \tag{1}$$

In this document, we define $\mathfrak{e}_{ijk} = \mathring{\mathfrak{e}}_i * \dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^*$ and the projection

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{Z}}) = \sum_{ijk}\delta_{ijk}z_{ijk}\mathfrak{e}_{ijk},$$

where $\delta_{ijk} = 1_{(i,j,k)\in\boldsymbol{\Omega}}$, where $1_{(\cdot)}$ is the indicator function. Also $\boldsymbol{\Omega}^c$ denotes the complement of $\boldsymbol{\Omega}$ and $\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}$ is the projection onto $\boldsymbol{\Omega}^c$. Denote $\boldsymbol{T}$ by the set

$$\boldsymbol{T} = \{\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{Y}}^* + \boldsymbol{\mathcal{W}} * \boldsymbol{\mathcal{V}}^*,\ \boldsymbol{\mathcal{Y}},\boldsymbol{\mathcal{W}} \in \mathbb{R}^{n\times r\times n_3}\}, \tag{2}$$

and by $\boldsymbol{T}^\perp$ its orthogonal complement. Then the projections onto $\boldsymbol{T}$ and $\boldsymbol{T}^\perp$ are respectively

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{Z}}) = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^* * \boldsymbol{\mathcal{Z}} + \boldsymbol{\mathcal{Z}} * \boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{V}}^* - \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^* * \boldsymbol{\mathcal{Z}} * \boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{V}}^*,$$

$$\begin{aligned}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}(\boldsymbol{\mathcal{Z}}) &= \boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{Z}})\\ &= (\boldsymbol{\mathcal{I}}_{n_1} - \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^*) * \boldsymbol{\mathcal{Z}} * (\boldsymbol{\mathcal{I}}_{n_2} - \boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{V}}^*),\end{aligned}$$

where $\boldsymbol{\mathcal{I}}_n$ denotes the $n\times n\times n_3$ identity tensor. Note that $\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}$ is self-adjoint. So we have

$$\begin{aligned}&\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})\|_F^2\\ =&\langle\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk}),\mathfrak{e}_{ijk}\rangle\\ =&\langle\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^* * \mathfrak{e}_{ijk} + \mathfrak{e}_{ijk} * \boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{V}}^* - \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^* * \mathfrak{e}_{ijk} * \boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{V}}^*, \mathfrak{e}_{ijk}\rangle.\end{aligned}$$

Note that

$$\begin{aligned}&\langle\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^* * \mathfrak{e}_{ijk},\mathfrak{e}_{ijk}\rangle\\ =&\left\langle\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^* * \mathring{\mathfrak{e}}_i * \dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^*, \mathring{\mathfrak{e}}_i * \dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^*\right\rangle\\ =&\left\langle\boldsymbol{\mathcal{U}}^* * \mathring{\mathfrak{e}}_i, \boldsymbol{\mathcal{U}}^* * \mathring{\mathfrak{e}}_i * (\dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^* * \mathring{\mathfrak{e}}_j * \dot{\mathfrak{e}}_k^*)\right\rangle\\ =&\langle\boldsymbol{\mathcal{U}}^* * \mathring{\mathfrak{e}}_i, \boldsymbol{\mathcal{U}}^* * \mathring{\mathfrak{e}}_i\rangle\\ =&\|\boldsymbol{\mathcal{U}}^* * \mathring{\mathfrak{e}}_i\|_F^2,\end{aligned}$$

where we use the fact that $\dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^* * \mathring{\mathfrak{e}}_j * \dot{\mathfrak{e}}_k^* = \boldsymbol{\mathcal{I}}_1$, which is the $1\times 1\times n_3$ identity tensor. Therefore, it is easy to see that

$$\begin{aligned}&\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})\|_F^2\\ =&\|\boldsymbol{\mathcal{U}}^* * \mathring{\mathfrak{e}}_i\|_F^2 + \|\boldsymbol{\mathcal{V}}^* * \mathring{\mathfrak{e}}_j\|_F^2 - \|\boldsymbol{\mathcal{U}}^* * \mathring{\mathfrak{e}}_i * \dot{\mathfrak{e}}_k * \mathring{\mathfrak{e}}_j^* * \boldsymbol{\mathcal{V}}\|_F^2,\\ \leq&\|\boldsymbol{\mathcal{U}}^* * \mathring{\mathfrak{e}}_i\|_F^2 + \|\boldsymbol{\mathcal{V}}^* * \mathring{\mathfrak{e}}_j\|_F^2\\ \leq&\frac{\mu r(n_1+n_2)}{n_1n_2n_3} &(3)\\ =&\frac{2\mu r}{nn_3},\ \text{when } n_1=n_2=n. &(4)\end{aligned}$$

*Corresponding author.

The following Tensor Incoherence Conditions will be used in the proofs

$$\max_{i=1,\cdots,n_1} \|\boldsymbol{\mathcal{U}}^* * \mathring{\mathfrak{e}}_i\|_F \le \sqrt{\frac{\mu r}{n_1 n_3}}, \tag{5}$$

$$\max_{j=1,\cdots,n_2} \|\boldsymbol{\mathcal{V}}^* * \mathring{\mathfrak{e}}_j\|_F \le \sqrt{\frac{\mu r}{n_2 n_3}}, \tag{6}$$

$$\|\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^*\|_\infty \le \sqrt{\frac{\mu r}{n_1 n_2 n_3^2}}. \tag{7}$$

3 Dual Certification

Theorem 3.1. ***(Subgradient of tensor nuclear norm)*** *Let $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ with $\text{rank}_t(\boldsymbol{\mathcal{A}}) = r$ and its skinny t-SVD be $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*$. Denote $\partial\|\boldsymbol{\mathcal{A}}\|_*$ be the set of the subgradients of the tensor nuclear norm at $\boldsymbol{\mathcal{A}}$. We have $\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{W}} \in \partial\|\boldsymbol{\mathcal{A}}\|_*$, where $\boldsymbol{\mathcal{W}}$ satisfies $\boldsymbol{\mathcal{U}}^* * \boldsymbol{\mathcal{W}} = \mathbf{0}$, $\boldsymbol{\mathcal{W}} * \boldsymbol{\mathcal{V}} = \mathbf{0}$ and $\|\boldsymbol{\mathcal{W}}\| \le 1$.*

Proof. Let $\boldsymbol{\mathcal{G}} \in \partial\|\boldsymbol{\mathcal{A}}\|_*$. It is equivalent to the following statements [4]

$$\|\boldsymbol{\mathcal{A}}\|_* = \langle \boldsymbol{\mathcal{G}}, \boldsymbol{\mathcal{A}} \rangle, \tag{8}$$

$$\|\boldsymbol{\mathcal{G}}\| \le 1. \tag{9}$$

So, to complete the proof, we only need to show that $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{W}}$, where $\boldsymbol{\mathcal{U}}^* * \boldsymbol{\mathcal{W}} = \mathbf{0}$, $\boldsymbol{\mathcal{W}} * \boldsymbol{\mathcal{V}} = \mathbf{0}$ and $\|\boldsymbol{\mathcal{W}}\| \le 1$, satisfies (8) and (9). First, we have

$$\begin{aligned}\langle \boldsymbol{\mathcal{G}}, \boldsymbol{\mathcal{A}} \rangle &= \langle \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{W}}, \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^* \rangle \\ &= \langle \boldsymbol{\mathcal{I}}, \boldsymbol{\mathcal{S}} \rangle + 0 = \frac{1}{n_3} \langle \bar{\boldsymbol{I}}, \bar{\boldsymbol{S}} \rangle \\ &= \frac{1}{n_3} \|\bar{\boldsymbol{A}}\|_* = \|\boldsymbol{\mathcal{A}}\|_*.\end{aligned}$$

Also, (9) is obvious when considering the property of $\boldsymbol{\mathcal{W}}$. □

3.1 Dual Certificates

Lemma 3.2. *Assume that $\|\mathcal{P}_{\boldsymbol{\Omega}}\mathcal{P}_{\boldsymbol{T}}\| < 1$. Then $(\boldsymbol{\mathcal{L}}_0, \boldsymbol{\mathcal{S}}_0)$ is the unique solution to the TRPCA problem if there is a pair $(\boldsymbol{\mathcal{W}}, \boldsymbol{F})$ obeying*

$$\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{W}} = \lambda(\text{sgn}\,(\boldsymbol{\mathcal{S}}_0) + \boldsymbol{F}),$$

with $\mathcal{P}_{\boldsymbol{T}}\boldsymbol{\mathcal{W}} = \mathbf{0}$, $\|\boldsymbol{\mathcal{W}}\| < 1$, $\mathcal{P}_{\boldsymbol{\Omega}}\boldsymbol{F} = \mathbf{0}$ and $\|\boldsymbol{F}\|_\infty < 1$.

Lemma 3.3. *Assume that $\|\mathcal{P}_{\boldsymbol{\Omega}}\mathcal{P}_{\boldsymbol{T}}\| \le \frac{1}{2}$ and $\lambda < \frac{1}{\sqrt{n_3}}$. Then $(\boldsymbol{\mathcal{L}}_0, \boldsymbol{\mathcal{S}}_0)$ is the unique solution to the TRPCA problem if there is a pair $(\boldsymbol{\mathcal{W}}, \boldsymbol{F})$ obeying*

$$(\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{W}}) = \lambda(\text{sgn}\,(\boldsymbol{\mathcal{S}}_0) + \boldsymbol{F} + \mathcal{P}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{D}}),$$

with $\mathcal{P}_{\boldsymbol{T}}\boldsymbol{\mathcal{W}} = \mathbf{0}$, $\|\boldsymbol{\mathcal{W}}\| \le \frac{1}{2}$, $\mathcal{P}_{\boldsymbol{\Omega}}\boldsymbol{F} = \mathbf{0}$ and $\|\boldsymbol{F}\|_\infty \le \frac{1}{2}$, and $\|\mathcal{P}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{D}}\|_F \le \frac{1}{4}$.

Lemma 3.3 implies that it is suffices to produce a dual certificate $\boldsymbol{\mathcal{W}}$ obeying

$$\begin{cases} \boldsymbol{\mathcal{W}} \in \boldsymbol{T}^\perp, \\ \|\boldsymbol{\mathcal{W}}\| < \frac{1}{2}, \\ \|\mathcal{P}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{W}} - \lambda\,\text{sgn}\,(\boldsymbol{\mathcal{S}}_0))\|_F \le \frac{\lambda}{4}, \\ \|\mathcal{P}_{\boldsymbol{\Omega}^\perp}(\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{W}})\|_\infty < \frac{\lambda}{2}. \end{cases} \tag{10}$$

3.2 Dual Certification via the Golfing Scheme

Before we introduce our construction, our model assumes that $\boldsymbol{\Omega} \sim \text{Ber}(\rho)$, or equivalently that $\boldsymbol{\Omega}^c \sim \text{Ber}(1-\rho)$. Now the distribution of $\boldsymbol{\Omega}^c$ is the same as that of $\boldsymbol{\Omega}^c = \boldsymbol{\Omega}_1 \cup \boldsymbol{\Omega}_2 \cup \cdots \cup \boldsymbol{\Omega}_{j_0}$, where each $\boldsymbol{\Omega}_j$ follows the Bernoulli model with parameter q, which satisfies

$$\mathbb{P}((i,j,k) \in \boldsymbol{\Omega}) = \mathbb{P}(\text{Bin}(j_0, q) = 0) = (1-q)^{j_0}, \tag{11}$$

so that the two models are the same if $\rho = (1-q)^{j_0}$. Note that because of overlaps between the $\boldsymbol{\Omega}_j$'s, $q \ge (1-\rho)/j_0$.

Now, we construct a dual certificate

$$\boldsymbol{\mathcal{W}} = \boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{L}}} + \boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}, \tag{12}$$

where each component is as follows:

1. Construction of $\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{L}}}$ via the Golfing Scheme. Let $j_0 = 2\log(nn_3)$ and $\boldsymbol{\Omega}_j$, $j = 1, \cdots, j_0$, be defined as previously described so that $\boldsymbol{\Omega}^c = \cup_{1\le j\le j_0}\boldsymbol{\Omega}_j$. Then define

$$\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{L}}} = \mathcal{P}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{Y}}_{j_0}, \tag{13}$$

   where

$$\boldsymbol{\mathcal{Y}}_j = \boldsymbol{\mathcal{Y}}_{j-1} + q^{-1}\mathcal{P}_{\boldsymbol{\Omega}_j}\mathcal{P}_{\boldsymbol{T}}(\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* - \boldsymbol{\mathcal{Y}}_{j-1}),\ \boldsymbol{\mathcal{Y}}_0 = \mathbf{0}.$$

2. Construction of $\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}$ via the Method of Least Squares. Assume that $\|\mathcal{P}_{\boldsymbol{\Omega}}\mathcal{P}_{\boldsymbol{T}}\| < 1/2$. Then, $\|\mathcal{P}_{\boldsymbol{\Omega}}\mathcal{P}_{\boldsymbol{T}}\mathcal{P}_{\boldsymbol{\Omega}}\| < 1/4$, and thus, the operator $\mathcal{P}_{\boldsymbol{\Omega}} - \mathcal{P}_{\boldsymbol{\Omega}}\mathcal{P}_{\boldsymbol{T}}\mathcal{P}_{\boldsymbol{\Omega}}$ mapping $\boldsymbol{\Omega}$ onto itself is invertible; we denote its inverse by $(\mathcal{P}_{\boldsymbol{\Omega}} - \mathcal{P}_{\boldsymbol{\Omega}}\mathcal{P}_{\boldsymbol{T}}\mathcal{P}_{\boldsymbol{\Omega}})^{-1}$. We then set

$$\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}} = \lambda\mathcal{P}_{\boldsymbol{T}^\perp}(\mathcal{P}_{\boldsymbol{\Omega}} - \mathcal{P}_{\boldsymbol{\Omega}}\mathcal{P}_{\boldsymbol{T}}\mathcal{P}_{\boldsymbol{\Omega}})^{-1}\,\text{sgn}\,(\boldsymbol{\mathcal{S}}_0). \tag{14}$$

   This is equivalent to

$$\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}} = \lambda\mathcal{P}_{\boldsymbol{T}^\perp}\sum_{k\ge 0}(\mathcal{P}_{\boldsymbol{\Omega}}\mathcal{P}_{\boldsymbol{T}}\mathcal{P}_{\boldsymbol{\Omega}})^k\,\text{sgn}\,(\boldsymbol{\mathcal{S}}_0). \tag{15}$$

Since both $\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{L}}}$ and $\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}$ belong to $\boldsymbol{T}^\perp$ and $\mathcal{P}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}} = \lambda\mathcal{P}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{I}} - \mathcal{P}_{\boldsymbol{T}})(\mathcal{P}_{\boldsymbol{\Omega}} - \mathcal{P}_{\boldsymbol{\Omega}}\mathcal{P}_{\boldsymbol{T}}\mathcal{P}_{\boldsymbol{\Omega}})^{-1}\,\text{sgn}\,(\boldsymbol{\mathcal{S}}_0) = \lambda\,\text{sgn}\,(\boldsymbol{\mathcal{S}}_0)$, we will establish that $\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{L}}} + \boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}$ is a valid dual certificate if it obeys

$$\begin{cases} \|\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{L}}} + \boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}\| < \frac{1}{2}, \\ \|\mathcal{P}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{L}}})\|_F \le \frac{\lambda}{4}, \\ \|\mathcal{P}_{\boldsymbol{\Omega}^\perp}(\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^* + \boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{L}}} + \boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}\|_\infty < \frac{\lambda}{2}. \end{cases} \tag{16}$$

This can be done by using the following two key lemmas:

Lemma 3.4. *Assume that $\mathbf{\Omega} \sim Ber(\rho)$ with parameter $\rho \le \rho_s$ for some $\rho_s > 0$. Set $j_0 = 2\lceil \log(nn_3) \rceil$ (use $\log(n_{(1)}n_3)$ for the tensors of rectangular frontal slice). Then, the tensor $\mathcal{W}^{\mathcal{L}}$ obeys*

(a) $\|\mathcal{W}^{\mathcal{L}}\| < \frac{1}{4}$,

(b) $\|\mathcal{P}_{\mathbf{\Omega}}(\mathcal{U} * \mathcal{V}^* + \mathcal{W}^{\mathcal{L}})\|_F < \frac{\lambda}{4}$,

(c) $\|\mathcal{P}_{\mathbf{\Omega}^\perp}(\mathcal{U} * \mathcal{V}^* + \mathcal{W}^{\mathcal{L}})\|_\infty < \frac{\lambda}{4}$.

Lemma 3.5. *Assume that $\mathcal{S}_0$ is supported on a set $\mathbf{\Omega}$ sampled as in Lemma 3.4, and that the signs of $\mathcal{S}_0$ are independent and identically distributed symmetric (and independent of $\mathbf{\Omega}$). Then, the tensor $\mathcal{W}^{\mathcal{S}}$ (14) obeys*

(a) $\|\mathcal{W}^{\mathcal{S}}\| < \frac{1}{4}$,

(b) $\|\mathcal{P}_{\mathbf{\Omega}^\perp}\mathcal{W}^{\mathcal{S}}\|_\infty < \frac{\lambda}{4}$.

4 Proofs of Dual Certification

4.1 Preliminaries

Lemma 4.1. *Suppose $\mathbf{\Omega} \sim Ber(\rho)$. Then with high probability,*

$$\|\mathcal{P}_{\boldsymbol{T}} - \rho^{-1}\mathcal{P}_{\boldsymbol{T}}\mathcal{P}_{\mathbf{\Omega}}\mathcal{P}_{\boldsymbol{T}}\| \le \epsilon, \tag{17}$$

provided that $\rho \ge C_0\epsilon^{-2}(\mu r \log(nn_3))/(nn_3)$ for some numerical constant $C_0 > 0$. For the tensor of rectangular frontal slice, we need $\rho \ge C_0\epsilon^{-2}(\mu r \log(n_{(1)}n_3))/(n_{(2)}n_3)$.

Corollary 4.2. *Assume that $\mathbf{\Omega} \sim Ber(\rho)$, then $\|\mathcal{P}_{\mathbf{\Omega}}\mathcal{P}_{\boldsymbol{T}}\|^2 \le \rho + \epsilon$, provided that $1-\rho \ge C\epsilon^{-2}(\mu r \log(nn_3))/(nn_3)$, where C is as in Lemma 4.1. For the tensor with frontal slice, the modification is as in Lemma 4.1.*

Note that this corollary shows that $\|\mathcal{P}_{\mathbf{\Omega}}\mathcal{P}_{\boldsymbol{T}}\| \le 1/2$, provided $|\mathbf{\Omega}|$ is not too large.

Lemma 4.3. *Suppose that $\mathcal{Z} \in \boldsymbol{T}$ is a fixed tensor, and $\mathbf{\Omega} \sim Ber(\rho)$. Then, with high probability,*

$$\|\mathcal{Z} - \rho^{-1}\mathcal{P}_{\boldsymbol{T}}\mathcal{P}_{\mathbf{\Omega}}\mathcal{Z}\|_\infty \le \epsilon\|\mathcal{Z}\|_\infty, \tag{18}$$

provided that $\rho \ge C_0\epsilon^{-2}(\mu r \log(nn_3))/(nn_3)$ (for the tensor of rectangular frontal slice, $\rho \ge C_0\epsilon^{-2}(\mu r \log(n_{(1)}n_3))/(n_{(2)}n_3)$) for some numerical constant $C_0 > 0$.

Lemma 4.4. *Suppose $\mathcal{Z}$ is fixed, and $\mathbf{\Omega} \sim Ber(\rho)$. Then, with high probability,*

$$\|(\mathcal{I} - \rho^{-1}\mathcal{P}_{\mathbf{\Omega}})\mathcal{Z}\| \le \sqrt{\frac{C_0 nn_3 \log(nn_3)}{\rho}}\|\mathcal{Z}\|_\infty, \tag{19}$$

for some numerical constant $C_0 > 0$ provided that $\rho \ge C_0 \log(nn_3)/(nn_3)$ (or $\rho \ge C_0 \log(n_{(1)}n_3)/(n_{(2)}n_3)$ for the tensors with rectangular frontal slice).

4.2 Proof of Lemma 3.4

Proof. We first introduce some notations. Set $\mathcal{Z}_j = \mathcal{U} * \mathcal{V}^* - \mathcal{P}_{\boldsymbol{T}}\mathcal{Y}_j$ obeying

$$\mathcal{Z}_j = (\mathcal{P}_{\boldsymbol{T}} - q^{-1}\mathcal{P}_{\boldsymbol{T}}\mathcal{P}_{\mathbf{\Omega}_j}\mathcal{P}_{\boldsymbol{T}})\mathcal{Z}_{j-1}.$$

So $\mathcal{Z}_j \in \mathcal{T}$ for all $j \ge 0$. Also, note that when

$$q \ge C_0\epsilon^{-2}\frac{\mu r \log(nn_3)}{nn_3}, \tag{20}$$

or for the tensors with rectangular frontal slices $q \ge C_0\epsilon^{-2}\frac{\mu r \log(n_{(1)}n_3)}{n_{(2)}n_3}$, we have

$$\|\mathcal{Z}_j\|_\infty \le \epsilon\|\mathcal{Z}_{j-1}\|_\infty \le \epsilon^j\|\mathcal{U} * \mathcal{V}^*\|_\infty, \tag{21}$$

by Lemma 4.3 and

$$\|\mathcal{Z}_j\|_F \le \epsilon\|\mathcal{Z}_{j-1}\|_F \le \epsilon^j\|\mathcal{U} * \mathcal{V}^*\|_F \le \epsilon^j\sqrt{r}. \tag{22}$$

We assume $\epsilon \le e^{-1}$.

1. Proof of (a). Note that $\mathcal{Y}_{j_0} = \sum_j q^{-1}\mathcal{P}_{\mathbf{\Omega}_j}\mathcal{Z}_{j-1}$. We have

$$\begin{aligned}
\|\mathcal{W}^{\mathcal{L}}\| =& \|\mathcal{P}_{\boldsymbol{T}^\perp}\mathcal{Y}_{j_0}\| \le \sum_j \|q^{-1}\mathcal{P}_{\boldsymbol{T}^\perp}\mathcal{P}_{\mathbf{\Omega}_j}\mathcal{Z}_{j-1}\| \\
\le& \sum_j \|\mathcal{P}_{\boldsymbol{T}^\perp}(q^{-1}\mathcal{P}_{\mathbf{\Omega}_j}\mathcal{Z}_{j-1} - \mathcal{Z}_{j-1})\| \\
\le& \sum_j \|q^{-1}\mathcal{P}_{\mathbf{\Omega}_j}\mathcal{Z}_{j-1} - \mathcal{Z}_{j-1}\| \\
\le& C_0'\sqrt{\frac{nn_3\log(nn_3)}{q}}\sum_j\|\mathcal{Z}_{j-1}\|_\infty \\
\le& C_0'\sqrt{\frac{nn_3\log(nn_3)}{q}}\sum_j\epsilon^{j-1}\|\mathcal{U} * \mathcal{V}^*\|_\infty \\
\le& C_0'(1-\epsilon)^{-1}\sqrt{\frac{nn_3\log(nn_3)}{q}}\|\mathcal{U} * \mathcal{V}^*\|_\infty.
\end{aligned}$$

The fourth step is from Lemma 4.4 and the fifth is from (21). Now by using (20) and (7), we have

$$\|\mathcal{W}^{\mathcal{L}}\| \le C'\epsilon,$$

for some numerical constant C'.

2. Proof of (b). Since $\mathcal{P}_{\mathbf{\Omega}}\mathcal{Y}_{j_0} = \mathbf{0}$, $\mathcal{P}_{\mathbf{\Omega}}(\mathcal{U} * \mathcal{V}^* + \mathcal{P}_{\boldsymbol{T}^\perp}\mathcal{Y}_{j_0}) = \mathcal{P}_{\mathbf{\Omega}}(\mathcal{U} * \mathcal{V}^* - \mathcal{P}_{\boldsymbol{T}}\mathcal{Y}_{j_0}) = \mathcal{P}_{\mathbf{\Omega}}(\mathcal{Z}_{j_0})$, and it follows from (22) that

$$\|\mathcal{Z}_{j_0}\|_F \le \epsilon^{j_0}\|\mathcal{U} * \mathcal{V}^*\|_F \le \epsilon^{j_0}\sqrt{r}.$$

Since $\epsilon \le e^{-1}$ and $j_0 \ge 2\log(nn_3)$, $\epsilon^{j_0} \le (nn_3)^{-2}$ and this proves the claim.

3. Proof of (c). We have $\mathcal{U} * \mathcal{V}^* + \mathcal{W}^{\mathcal{L}} = \mathcal{Z}_{j_0} + \mathcal{Y}_{j_0}$ and know that $\mathcal{Y}_{j_0}$ is supported on $\mathbf{\Omega}^c$. Therefore, since $\|\mathcal{Z}_{j_0}\|_F \le \lambda/8$.

We only need to show that $\|\boldsymbol{\mathcal{Y}}_{j_0}\|_\infty \leq \lambda/8$. Indeed,

$$\begin{aligned}\|\boldsymbol{\mathcal{Y}}_{j_0}\|_\infty \leq & q^{-1}\sum_j \|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}_j}\boldsymbol{\mathcal{Z}}_{j-1}\|_\infty \\ \leq & q^{-1}\sum_j \|\boldsymbol{\mathcal{Z}}_{j-1}\|_\infty \\ \leq & q^{-1}\sum_j \epsilon^{j-1}\|\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^*\|_\infty.\end{aligned}$$

Since $\|\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{V}}^*\|_\infty \leq \sqrt{\frac{\mu r}{n^2 n_3^2}}$, this gives

$$\|\boldsymbol{\mathcal{Y}}_{j_0}\|_\infty \leq C' \frac{\epsilon^2}{\sqrt{\mu r (\log(nn_3))^2}},$$

for some numerical constant C' whenever q obeys (20). Since $\lambda = 1/\sqrt{nn_3}$, $\|\boldsymbol{\mathcal{Y}}_{j_0}\|_\infty \leq \lambda/8$ if

$$\epsilon \leq C\left(\frac{\mu r(\log(nn_3))^2}{nn_3}\right)^{1/4}.$$

The claim is proved by using (20), (7) and sufficiently small ϵ (provided that ρ_r is sufficiently small. Note that everything is consistent since $C_0\epsilon^{-2}\frac{\mu r \log(nn_3)}{nn_3} < 1$. □

4.3 Proof of Lemma 3.5

Proof. We denote $\boldsymbol{\mathcal{M}} = \operatorname{sgn}(\boldsymbol{\mathcal{S}}_0)$ distributed as

$$\boldsymbol{\mathcal{M}}_{ijk} = \begin{cases} 1, & \text{w.p. } \rho/2, \\ 0, & \text{w.p. } 1-\rho, \\ -1, & \text{w.p. } \rho/2. \end{cases} \tag{23}$$

Note that for any $\sigma > 0$, $\{\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}\| \leq \sigma\}$ holds with high probability provided that ρ is sufficiently small, see Corollary 4.2.

1. Proof of (a). By construction,

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

Note that $\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{W}}_0^{\boldsymbol{\mathcal{S}}}\| \leq \|\boldsymbol{\mathcal{W}}_0^{\boldsymbol{\mathcal{S}}}\| = \lambda\|\boldsymbol{\mathcal{M}}\|$ and $\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{W}}_1^{\boldsymbol{\mathcal{S}}}\| \leq \|\boldsymbol{\mathcal{W}}_1^{\boldsymbol{\mathcal{S}}}\| = \lambda\|\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})\|$, where $\boldsymbol{\mathcal{R}} = \sum_{k\geq 1}(\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}})^k$. Now, we will respectively show that $\lambda\|\boldsymbol{\mathcal{M}}\|$ and $\lambda\|\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})\|$ are small enough when ρ is sufficiently small for $\lambda = 1/\sqrt{nn_3}$. Therefor, $\|\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}\| \leq 1/4$.

1) Bound $\|\boldsymbol{\mathcal{M}}\|$.

Lemma 4.5. *For the Bernoulli sign variable $\boldsymbol{\mathcal{M}}$ defined in (23), there exists a function $\varphi(\rho)$ satisfying $\lim_{\rho\to 0^+}\varphi(\rho) = 0$, such that the following statement holds with with large probability,*

$$\|\boldsymbol{\mathcal{M}}\| \leq \varphi(\rho)\sqrt{nn_3}. \tag{24}$$

The proof has three steps.

Step 1: Approximation. We first introduce some notations. Let $\boldsymbol{f}_i^*$ be the i-th row of $\boldsymbol{F}_{n_3}$, and $\boldsymbol{M}^H = \begin{bmatrix}\boldsymbol{M}_1^H \\ \boldsymbol{M}_2^H \\ \vdots \\ \boldsymbol{M}_n^H\end{bmatrix} \in \mathbb{R}^{nn_3\times n}$ be a matrix unfolded by $\boldsymbol{\mathcal{M}}$, where $\boldsymbol{M}_i^H \in \mathbb{R}^{n_3\times n}$ is the i-th horizontal slice of $\boldsymbol{\mathcal{M}}$, i.e., $[\boldsymbol{M}_i^H]_{kj} = \boldsymbol{\mathcal{M}}_{ikj}$. Consider that $\bar{\boldsymbol{\mathcal{M}}} = \texttt{fft}(\boldsymbol{\mathcal{M}}, [], 3)$, we have

$$\bar{\boldsymbol{M}}_i = \begin{bmatrix}\boldsymbol{f}_i^*\boldsymbol{M}_1^H \\ \boldsymbol{f}_i^*\boldsymbol{M}_2^H \\ \vdots \\ \boldsymbol{f}_i^*\boldsymbol{M}_n^H\end{bmatrix}, \tag{25}$$

where $\bar{\boldsymbol{M}}_i \in \mathbb{R}^{n\times n}$ is the i-th frontal slice of $\boldsymbol{\mathcal{M}}$. Note that

$$\|\boldsymbol{\mathcal{M}}\| = \|\bar{\boldsymbol{M}}\| = \max_{i=1,\cdots,n_3}\|\bar{\boldsymbol{M}}_i\|. \tag{26}$$

Let N be the $1/2$-net for $\mathbb{S}^{n-1}$ of size at most 5^n (see Lemma 5.2 in [3]). Then Lemma 5.3 in [3] gives

$$\|\bar{\boldsymbol{M}}_i\| \leq 2\max_{\boldsymbol{x}\in N}\|\bar{\boldsymbol{M}}_i\boldsymbol{x}\|_2. \tag{27}$$

So we consider to bound $\|\bar{\boldsymbol{M}}_i\boldsymbol{x}\|_2$.

Step 2: Concentration. We can express $\|\bar{\boldsymbol{M}}_i\boldsymbol{x}\|_2^2$ as a sum of independent random variables

$$\|\bar{\boldsymbol{M}}_i\boldsymbol{x}\|_2^2 = \sum_{j=1}^n(\boldsymbol{f}_i^*\boldsymbol{M}_j^H\boldsymbol{x})^2 := \sum_{j=1}^n z_j^2, \tag{28}$$

where $z_j = \langle\boldsymbol{M}_j^H, \boldsymbol{f}_i\boldsymbol{x}^*\rangle$, $j = 1,\cdots,n$, are independent sub-gaussian random variables with $\mathbb{E}z_j^2 = \rho\|\boldsymbol{f}_i\boldsymbol{x}^*\|_F^2 = \rho n_3$. Using (23), we have

$$|[\boldsymbol{M}_j^H]_{kl}| = \begin{cases}1, & \text{w.p. } \rho, \\ 0, & \text{w.p. } 1-\rho.\end{cases}$$

Thus, the sub-gaussian norm of $[\boldsymbol{M}_j^H]_{kl}$, denoted as $\|\cdot\|_{\psi_2}$, is

$$\begin{aligned}\|[\boldsymbol{M}_j^H]_{kl}\|_{\psi_2} = & \sup_{p\geq 1} p^{-\frac{1}{2}}(\mathbb{E}[|[\boldsymbol{M}_j^H]_{kl}|^p])^{\frac{1}{p}} \\ = & \sup_{p\geq 1} p^{-\frac{1}{2}}\rho^{\frac{1}{p}}.\end{aligned}$$

Define the function $\phi(x) = x^{-\frac{1}{2}}\rho^{\frac{1}{x}}$ on $[1,+\infty)$. The only stationary point occurs at $x^* = \log\rho^{-2}$. Thus,

$$\begin{aligned}\phi(x) \leq & \max(\phi(1),\phi(x^*)) \\ = & \max\left(\rho, (\log\rho^{-2})^{-\frac{1}{2}}\rho^{\frac{1}{\log\rho^{-2}}}\right) \\ := & \psi(\rho).\end{aligned} \tag{29}$$

Therefore, $\|[\boldsymbol{M}_j^H]_{kl}\|_{\psi_2} \leq \psi(\rho)$. Consider that z_j is a sum of independent centered sub-gaussian random variables $[\boldsymbol{M}_j^H]_{kl}$'s, by using Lemma 5.9 in [3], we have $\|z_j\|_{\psi_2}^2 \leq c_1(\psi(\rho))^2 n_3$, where c_1 is an absolute constant. Therefore, by Remark 5.18 and Lemma 5.14 in [3], $z_j^2 - \rho n_3$ are independent centered sub-exponential random variables with $\|z_j^2 - \rho n_3\|_{\psi_1} \leq 2\|z_j\|_{\psi_1}^2 \leq 4\|z_j\|_{\psi_2}^2 \leq 4c_1(\psi(\rho))^2 n_3$.

Now, we use an exponential deviation inequality, Corollary 5.17 in [3], to control the sum (28). We have

$$
\begin{aligned}
&\mathbb{P}\left(\left|\|\bar{\boldsymbol{M}}_i\boldsymbol{x}\|_2^2 - \rho n n_3\right| \geq tn\right) \\
=&\mathbb{P}\left(\left|\sum_{j=1}^{n}(z_j^2 - \rho n_3)\right| \geq tn\right) \\
\leq&2\exp\left(-c_2 n \min\left(\left(\frac{t}{4c_1(\psi(\rho))^2 n_3}\right)^2, \frac{t}{4c_1(\psi(\rho))^2 n_3}\right)\right),
\end{aligned}
$$

where $c_2 > 0$. Let $t = c_3(\psi(\rho))^2 n_3$ for some absolute constant c_3, we have

$$
\begin{aligned}
&\mathbb{P}\left(\left|\|\bar{\boldsymbol{M}}_i\boldsymbol{x}\|_2^2 - \rho n n_3\right| \geq c_3(\psi(\rho))^2 n n_3\right) \\
\leq&2\exp\left(-c_2 n \min\left(\left(\frac{c_3}{4c_1}\right)^2, \frac{c_3}{4c_1}\right)\right).
\end{aligned}
$$

Step 3 Union bound. Taking the union bound over all $\boldsymbol{x}$ in the net N of cardinality $|N| \leq 5^n$, we obtain

$$
\begin{aligned}
&\mathbb{P}\left(\left|\max_{\boldsymbol{x}\in N}\|\bar{\boldsymbol{M}}_i\boldsymbol{x}\|_2^2 - \rho n n_3\right| \geq c_3(\psi(\rho))^2 n n_3\right) \\
\leq&2\cdot 5^n \cdot \exp\left(-c_2 n \min\left(\left(\frac{c_3}{4c_1}\right)^2, \frac{c_3}{4c_1}\right)\right).
\end{aligned}
$$

Furthermore, taking the union bound over all $i = 1, \cdots, n_3$, we have

$$
\begin{aligned}
&\mathbb{P}\left(\max_i \left|\max_{\boldsymbol{x}\in N}\|\bar{\boldsymbol{M}}_i\boldsymbol{x}\|_2^2 - \rho n n_3\right| \geq c_3(\psi(\rho))^2 n n_3\right) \\
\leq&2\cdot 5^n \cdot n_3 \cdot \exp\left(-c_2 n \min\left(\left(\frac{c_3}{4c_1}\right)^2, \frac{c_3}{4c_1}\right)\right).
\end{aligned}
$$

This implies that, with high probability (when the constant c_3 is large enough),

$$
\max_i \max_{\boldsymbol{x}\in N}\|\bar{\boldsymbol{M}}_i\boldsymbol{x}\|_2^2 \leq (\rho + c_3(\psi(\rho))^2) n n_3. \tag{30}
$$

Let $\varphi(\rho) = 2\sqrt{\rho + c_3(\psi(\rho))^2}$ and it satisfies $\lim_{\rho\to 0^+}\varphi(\rho) = 0$ by using (29). The proof is completed by further combining (26), (27) and (30).

2) Bound $\|\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})\|$.

For simplicity, let $\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})$. We have

$$
\|\boldsymbol{\mathcal{Z}}\| = \|\bar{\boldsymbol{Z}}\| = \sup_{\boldsymbol{x}\in\mathbb{S}^{nn_3-1}}\|\bar{\boldsymbol{Z}}\boldsymbol{x}\|_2. \tag{31}
$$

The optimal $\boldsymbol{x}$ to (31) is an eigenvector of $\bar{\boldsymbol{Z}}^*\bar{\boldsymbol{Z}}$. Since $\bar{\boldsymbol{Z}}$ is a block diagonal matrix, the optimal $\boldsymbol{x}$ has a block sparse structure, i.e., $\boldsymbol{x} \in B = \{\boldsymbol{x} \in \mathbb{R}^{nn_3} | \boldsymbol{x} = [\boldsymbol{x}_1^\top, \cdots, \boldsymbol{x}_i^\top \cdots, \boldsymbol{x}_{n_3}^\top]$, with $\boldsymbol{x}_i \in \mathbb{R}^n$, and there exists j such that $\boldsymbol{x}_j \neq \boldsymbol{0}$ and $\boldsymbol{x}_i = \boldsymbol{0}, i \neq j\}$. Note that $\|\boldsymbol{x}\|_2 = \|\boldsymbol{x}_j\|_2 = 1$. Let N be the $1/2$-net for $\mathbb{S}^{n-1}$ of size at most 5^n (see Lemma 5.2 in [3]). Then the $1/2$-net, denoted as N', for B has the size at most $n_3 \cdot 5^n$. We have

$$
\begin{aligned}
\|\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})\| =&\|\texttt{bdiag}(\overline{\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})})\| \\
=&\sup_{\boldsymbol{x},\boldsymbol{y}\in B}\left\langle \boldsymbol{x}, \texttt{bdiag}(\overline{\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})})\boldsymbol{y}\right\rangle \\
=&\sup_{\boldsymbol{x},\boldsymbol{y}\in B}\left\langle \boldsymbol{x}\boldsymbol{y}^*, \texttt{bdiag}(\overline{\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})})\right\rangle \\
=&\sup_{\boldsymbol{x},\boldsymbol{y}\in B}\left\langle \texttt{bdiag}^*(\boldsymbol{x}\boldsymbol{y}^*), \overline{\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})}\right\rangle,
\end{aligned}
$$

where $\texttt{bdiag}^*$, the joint operator of $\texttt{bdiag}$, maps the block diagonal matrix $\boldsymbol{x}\boldsymbol{y}^*$ to a tensor of size $n \times n \times n_3$. Let $\boldsymbol{\mathcal{Z}}' = \texttt{bdiag}^*(\boldsymbol{x}\boldsymbol{y}^*)$ and $\boldsymbol{\mathcal{Z}} = \texttt{ifft}(\boldsymbol{\mathcal{Z}}', [], 3)$. We have

$$
\begin{aligned}
\|\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})\| =& \sup_{\boldsymbol{x},\boldsymbol{y}\in B}\left\langle \boldsymbol{\mathcal{Z}}', \overline{\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})}\right\rangle \\
=& \sup_{\boldsymbol{x},\boldsymbol{y}\in B} n_3 \left\langle \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})\right\rangle \\
=& \sup_{\boldsymbol{x},\boldsymbol{y}\in B} n_3 \left\langle \boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{Z}}), \boldsymbol{\mathcal{M}}\right\rangle \\
\leq& \sup_{\boldsymbol{x},\boldsymbol{y}\in N'} 4n_3 \left\langle \boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{Z}}), \boldsymbol{\mathcal{M}}\right\rangle.
\end{aligned}
$$

For a fixed pair $(\boldsymbol{x}, \boldsymbol{y})$ of unit-normed vectors, define the random variable

$$
X(\boldsymbol{x}, \boldsymbol{y}) = \left\langle 4n_3\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{Z}}), \boldsymbol{\mathcal{M}}\right\rangle.
$$

Conditional on $\boldsymbol{\Omega} = \text{supp}(\boldsymbol{\mathcal{M}})$, the signs of $\boldsymbol{\mathcal{M}}$ are independent and identically distributed symmetric and Hoeffding's inequality gives

$$
\mathbb{P}(|X(\boldsymbol{x}, \boldsymbol{y})| > t|\boldsymbol{\Omega}) \leq 2\exp\left(\frac{-2t^2}{\|4n_3\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{Z}})\|_F^2}\right).
$$

Note that $\|4n_3\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{Z}}))\|_F \leq 4n_3\|\boldsymbol{\mathcal{R}}\|\|\boldsymbol{\mathcal{Z}}\|_F = 4\sqrt{n_3}\|\boldsymbol{\mathcal{R}}\|\|\boldsymbol{\mathcal{Z}}'\|_F = 4\sqrt{n_3}\|\boldsymbol{\mathcal{R}}\|$. Therefore, we have

$$
\mathbb{P}\left(\sup_{\boldsymbol{x},\boldsymbol{y}\in N'}|X(\boldsymbol{x}, \boldsymbol{y})| > t|\boldsymbol{\Omega}\right) \leq 2|N'|^2\exp\left(-\frac{t^2}{8n_3\|\boldsymbol{\mathcal{R}}\|^2}\right).
$$

Hence,

$$
\mathbb{P}\left(\|\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})\| > t|\boldsymbol{\Omega}\right) \leq 2|N'|^2\exp\left(-\frac{t^2}{8n_3\|\boldsymbol{\mathcal{R}}\|^2}\right).
$$

On the event $\{\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}\| \leq \sigma\}$,

$$
\|\boldsymbol{\mathcal{R}}\| \leq \sum_{k\geq 1}\sigma^{2k} = \frac{\sigma^2}{1-\sigma^2},
$$

and, therefore, unconditionally,

$$\begin{aligned}&\mathbb{P}\left(\|\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})\|>t\right)\\ \leq&2|N'|^2\exp\left(-\frac{\gamma^2t^2}{8n_3}\right)+\mathbb{P}\left(\|\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\|\geq\sigma\right),\ \gamma=\frac{1-\sigma^2}{2\sigma^2}\\ =&2n_3^2\cdot5^{2n}\exp\left(-\frac{\gamma^2t^2}{8n_3}\right)+\mathbb{P}\left(\|\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\|\geq\sigma\right).\end{aligned}$$

Let $t=c\sqrt{nn_3}$, where c can be a small absolute constant. Then the above inequality implies that $\|\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{M}})\|\leq t$ with high probability.

2. Proof of (b) Observe that

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}=-\lambda\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{P}_T}(\boldsymbol{\mathcal{P}_\Omega}-\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_\Omega})^{-1}\boldsymbol{\mathcal{M}}.$$

Now for $(i,j,k)\in\boldsymbol{\Omega}^c$, $\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}_{ijk}=\left\langle\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}},\mathfrak{e}_{ijk}\right\rangle$, and we have $\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}_{ijk}=\lambda\left\langle\boldsymbol{\mathcal{Q}}(i,j,k),\boldsymbol{\mathcal{M}}\right\rangle$, where $\boldsymbol{\mathcal{Q}}(i,j,k)$ is the tensor $-(\boldsymbol{\mathcal{P}_\Omega}-\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_\Omega})^{-1}\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}(\mathfrak{e}_{ijk})$. Conditional on $\boldsymbol{\Omega}=\mathrm{supp}(\boldsymbol{\mathcal{M}})$, the signs of $\boldsymbol{\mathcal{M}}$ are independent and identically distributed symmetric, and the Hoeffding's inequality gives

$$\mathbb{P}(|\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}_{ijk}|>t\lambda|\boldsymbol{\Omega})\leq2\exp\left(-\frac{2t^2}{\|\boldsymbol{\mathcal{Q}}(i,j,k)\|_F^2}\right),$$

and

$$\begin{aligned}&\mathbb{P}(\sup_{i,j,k}|\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}_{ijk}|>t\lambda/n_3|\boldsymbol{\Omega})\\ \leq&2n^2n_3\exp\left(-\frac{2t^2}{\sup_{i,j,k}\|\boldsymbol{\mathcal{Q}}(i,j,k)\|_F^2}\right).\end{aligned}$$

By using (4), we have

$$\begin{aligned}\|\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}(\mathfrak{e}_{ijk})\|_F\leq&\|\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\|\|\boldsymbol{\mathcal{P}_T}(\mathfrak{e}_{ijk})\|_F\\ \leq&\sigma\sqrt{\frac{2\mu r}{nn_3}},\end{aligned}$$

on the event $\{\|\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\|\leq\sigma\}$. On the same event, we have $\|(\boldsymbol{\mathcal{P}_\Omega}-\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_\Omega})^{-1}\|\leq(1-\sigma^2)^{-1}$ and thus $\|\boldsymbol{\mathcal{Q}}(i,j,k)\|_F^2\leq\frac{2\sigma^2}{(1-\sigma^2)^2}\frac{\mu r}{nn_3}$. Then, unconditionally,

$$\begin{aligned}&\mathbb{P}\left(\sup_{i,j,k}|\boldsymbol{\mathcal{W}}^{\boldsymbol{\mathcal{S}}}_{ijk}|>t\lambda\right)\\ \leq&2n^2n_3\exp\left(-\frac{nn_3\gamma^2t^2}{\mu r}\right)+\mathbb{P}(\|\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\|\geq\sigma),\end{aligned}$$

where $\gamma=\frac{(1-\sigma^2)^2}{2\sigma^2}$. This proves the claim when $\mu r<\rho_r'nn_3\log(nn_3)^{-1}$ and ρ_r' is sufficiently small. ☐

5 Proofs of Some Lemmas

Lemma 5.1. *[2] Consider a finite sequence $\{\boldsymbol{Z}_k\}$ of independent, random $n_1\times n_2$ matrices that satisfy the assumption $\mathbb{E}\boldsymbol{Z}_k=\boldsymbol{0}$ and $\|\boldsymbol{Z}_k\|\leq R$ almost surely. Let $\sigma^2=\max\{\|\sum_k\mathbb{E}[\boldsymbol{Z}_k\boldsymbol{Z}_k^*]\|,\max\{\|\sum_k\mathbb{E}[\boldsymbol{Z}_k^*\boldsymbol{Z}_k]\|\}$. Then, for any $t\geq0$, we have*

$$\mathbb{P}\left[\left\|\sum_k\boldsymbol{Z}_k\right\|\geq t\right]\leq(n_1+n_2)\exp\left(-\frac{t^2}{2\sigma^2+\frac{2}{3}Rt}\right)\tag{32}$$

$$\leq(n_1+n_2)\exp\left(-\frac{3t^2}{8\sigma^2}\right),\ \textit{for}\ t\leq\frac{\sigma^2}{R}.\tag{33}$$

Or, for any $c>0$, we have

$$\left\|\sum_k\boldsymbol{Z}_k\right\|\geq2\sqrt{c\sigma^2\log(n_1+n_2)}+cB\log(n_1+n_2),\tag{34}$$

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

5.1 Proof of Lemma 3.2

Proof. For any $\boldsymbol{\mathcal{H}}\neq\boldsymbol{0}$, $(\boldsymbol{\mathcal{L}}_0+\boldsymbol{\mathcal{H}},\boldsymbol{\mathcal{S}}_0-\boldsymbol{\mathcal{H}})$ is also a feasible solution. We show that its objective is larger than that at $(\boldsymbol{\mathcal{L}}_0,\boldsymbol{\mathcal{S}}_0)$, hence proving that $(\boldsymbol{\mathcal{L}}_0,\boldsymbol{\mathcal{S}}_0)$ is the unique solution. To do this, let $\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*+\boldsymbol{\mathcal{W}}_0$ be an arbitrary subgradient of the tensor nuclear norm at $\boldsymbol{\mathcal{L}}_0$, and $\mathrm{sgn}\,(\boldsymbol{\mathcal{S}}_0)+\boldsymbol{F}_0$ be an arbitrary subgradient of the ℓ_1-norm at $\boldsymbol{\mathcal{S}}_0$. Then we have

$$\begin{aligned}&\|\boldsymbol{\mathcal{L}}_0+\boldsymbol{\mathcal{H}}\|_*+\lambda\|\boldsymbol{\mathcal{S}}_0-\boldsymbol{\mathcal{H}}\|_1\\ \geq&\|\boldsymbol{\mathcal{L}}_0\|_*+\lambda\|\boldsymbol{\mathcal{S}}_0\|_1+\langle\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*+\boldsymbol{\mathcal{W}}_0,\boldsymbol{\mathcal{H}}\rangle\\ &-\lambda\left\langle\mathrm{sgn}\,(\boldsymbol{\mathcal{S}}_0)+\boldsymbol{F}_0,\boldsymbol{\mathcal{H}}\right\rangle.\end{aligned}$$

Now pick $\boldsymbol{\mathcal{W}}_0$ such that $\langle\boldsymbol{\mathcal{W}}_0,\boldsymbol{\mathcal{H}}\rangle=\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_*$ and $\langle\boldsymbol{F}_0,\boldsymbol{\mathcal{H}}\rangle=-\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{H}}\|$. We have

$$\begin{aligned}&\|\boldsymbol{\mathcal{L}}_0+\boldsymbol{\mathcal{H}}\|_*+\lambda\|\boldsymbol{\mathcal{S}}_0-\boldsymbol{\mathcal{H}}\|_1\\ \geq&\|\boldsymbol{\mathcal{L}}_0\|_*+\lambda\|\boldsymbol{\mathcal{S}}_0\|_1+\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_*+\lambda\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{H}}\|_1\\ &+\left\langle\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*-\lambda\,\mathrm{sgn}\,(\boldsymbol{\mathcal{S}}_0),\boldsymbol{\mathcal{H}}\right\rangle.\end{aligned}$$

By assumption

$$\begin{aligned}&\left|\left\langle\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}^*-\lambda\,\mathrm{sgn}\,(\boldsymbol{\mathcal{S}}_0),\boldsymbol{\mathcal{H}}\right\rangle\right|\\ \leq&\left|\langle\boldsymbol{\mathcal{W}},\boldsymbol{\mathcal{H}}\rangle\right|+\lambda\left|\langle\boldsymbol{F},\boldsymbol{\mathcal{H}}\rangle\right|\\ \leq&\beta(\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_*+\lambda\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{H}}\|_1),\end{aligned}$$

where $\beta=\max(\|\boldsymbol{\mathcal{W}}\|,\|\boldsymbol{F}\|_\infty)<1$. Thus

$$\begin{aligned}&\|\boldsymbol{\mathcal{L}}_0+\boldsymbol{\mathcal{H}}\|_*+\lambda\|\boldsymbol{\mathcal{S}}_0-\boldsymbol{\mathcal{H}}\|_1\\ \geq&\|\boldsymbol{\mathcal{L}}_0\|_*+\lambda\|\boldsymbol{\mathcal{S}}_0\|_1+(1-\beta)(\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_*+\lambda\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{H}}\|_1).\end{aligned}$$

Note that $\|\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\|<1$. This is equivalent to $\boldsymbol{\Omega}\cap\boldsymbol{T}=\{\boldsymbol{0}\}$. Thus $\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_*+\lambda\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{H}}\|_1>0$ unless $\boldsymbol{\mathcal{H}}=\boldsymbol{0}$. ☐

5.2 Proof of Lemma 3.3

Proof. Following the proof of Lemma 3.2, we have

$$
\begin{aligned}
&\|\boldsymbol{\mathcal{L}}_0+\boldsymbol{\mathcal{H}}\|_*+\lambda\|\boldsymbol{\mathcal{S}}_0-\boldsymbol{\mathcal{H}}\|_1\\
\geq&\|\boldsymbol{\mathcal{L}}_0\|_*+\lambda\|\boldsymbol{\mathcal{S}}_0\|_1+\frac{1}{2}(\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_*+\lambda\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{H}}\|_1)\\
&-\lambda\left\langle\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{D}},\boldsymbol{\mathcal{H}}\right\rangle\\
\geq&\|\boldsymbol{\mathcal{L}}_0\|_*+\lambda\|\boldsymbol{\mathcal{S}}_0\|_1+\frac{1}{2}(\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_*+\lambda\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{H}}\|_1)\\
&-\frac{\lambda}{4}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{H}}\|_F.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{H}}\|_F\leq&\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}\boldsymbol{\mathcal{H}}\|_F+\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_F\\
=&\frac{1}{2}\|\boldsymbol{\mathcal{H}}\|_F+\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_F\\
\leq&\frac{1}{2}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{H}}\|_F+\frac{1}{2}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{H}}\|_F+\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_F.
\end{aligned}
$$

Thus

$$
\begin{aligned}
\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{H}}\|_F\leq&\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{H}}\|_F+2\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_F\\
\leq&\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{H}}\|_1+2\sqrt{n_3}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_*.
\end{aligned}
$$

In conclusion,

$$
\begin{aligned}
&\|\boldsymbol{\mathcal{L}}_0+\boldsymbol{\mathcal{H}}\|_*+\lambda\|\boldsymbol{\mathcal{S}}_0-\boldsymbol{\mathcal{H}}\|_1\\
\geq&\|\boldsymbol{\mathcal{L}}_0\|_*+\lambda\|\boldsymbol{\mathcal{S}}_0\|_1+\frac{1}{2}\left(1-\lambda\sqrt{n_3}\right)\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^\perp}\boldsymbol{\mathcal{H}}\|_*\\
&+\frac{\lambda}{4}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}^\perp}\boldsymbol{\mathcal{H}}\|_1,
\end{aligned}
$$

where the last two terms are strictly positive when $\boldsymbol{\mathcal{H}}\neq\mathbf{0}$. □

5.3 Proof of Lemma 4.1

Proof. For any tensor $\boldsymbol{\mathcal{Z}}$, we can write

$$
\begin{aligned}
&(\rho^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}-\boldsymbol{\mathcal{P}}_{\boldsymbol{T}})\boldsymbol{\mathcal{Z}}\\
=&\sum_{ijk}\left(\rho^{-1}\delta_{ijk}-1\right)\left\langle\mathfrak{e}_{ijk},\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}\boldsymbol{\mathcal{Z}}\right\rangle\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})\\
:=&\sum_{ijk}\boldsymbol{\mathcal{H}}_{ijk}(\boldsymbol{\mathcal{Z}})
\end{aligned}
$$

where $\boldsymbol{\mathcal{H}}_{ijk}:\mathbb{R}^{n\times n\times n_3}\rightarrow\mathbb{R}^{n\times n\times n_3}$ is a self-adjoint random operator with $\mathbb{E}[\boldsymbol{\mathcal{H}}_{ijk}]=\mathbf{0}$. Define the matrix operator $\bar{\boldsymbol{H}}_{ijk}:\mathbb{B}\rightarrow\mathbb{B}$, where $\mathbb{B}=\{\bar{\boldsymbol{B}}:\boldsymbol{\mathcal{B}}\in\mathbb{R}^{n\times n\times n_3}\}$ denotes the set consists of block diagonal matrices with the blocks as the frontal slices of $\bar{\boldsymbol{\mathcal{B}}}$, as

$$
\bar{\boldsymbol{H}}_{ijk}(\bar{\boldsymbol{Z}})=\left(\rho^{-1}\delta_{ijk}-1\right)\left\langle\mathfrak{e}_{ijk},\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{Z}})\right\rangle\mathtt{bdiag}(\overline{\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})}).
$$

By the above definitions, we have $\|\boldsymbol{\mathcal{H}}_{ijk}\|=\|\bar{\boldsymbol{H}}_{ijk}\|$ and $\|\sum_{ijk}\boldsymbol{\mathcal{H}}_{ijk}\|=\|\sum_{ijk}\bar{\boldsymbol{H}}_{ijk}\|$. Also $\bar{\boldsymbol{H}}_{ijk}$ is self-adjoint and $\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}]=0$. To prove the result by the non-commutative Bernstein inequality, we need to bound $\|\bar{\boldsymbol{H}}_{ijk}\|$ and $\left\|\sum_{ijk}\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}^2]\right\|$. First, we have

$$
\begin{aligned}
\|\bar{\boldsymbol{H}}_{ijk}\|=&\sup_{\|\bar{\boldsymbol{Z}}\|_F=1}\|\bar{\boldsymbol{H}}_{ijk}(\bar{\boldsymbol{Z}})\|_F\\
\leq&\sup_{\|\bar{\boldsymbol{Z}}\|_F=1}\rho^{-1}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})\|_F\|\mathtt{bdiag}(\overline{\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})})\|_F\|\boldsymbol{\mathcal{Z}}\|_F\\
=&\sup_{\|\bar{\boldsymbol{Z}}\|_F=1}\rho^{-1}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})\|_F^2\|\bar{\boldsymbol{Z}}\|_F\\
\leq&\frac{2\mu r}{nn_3\rho},
\end{aligned}
$$

where the last inequality uses (4). On the other hand, by direct computation, we have $\bar{\boldsymbol{H}}_{ijk}^2(\bar{\boldsymbol{Z}})=(\rho^{-1}\delta_{ijk}-1)^2\left\langle\mathfrak{e}_{ijk},\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{Z}})\right\rangle\left\langle\mathfrak{e}_{ijk},\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})\right\rangle\mathtt{bdiag}(\overline{\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})})$. Note that $\mathbb{E}[(\rho^{-1}\delta_{ijk}-1)^2]\leq\rho^{-1}$. We have

$$
\begin{aligned}
&\left\|\sum_{ijk}\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}^2(\bar{\boldsymbol{Z}})]\right\|_F\\
\leq&\rho^{-1}\left\|\sum_{ijk}\left\langle\mathfrak{e}_{ijk},\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{Z}})\right\rangle\left\langle\mathfrak{e}_{ijk},\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})\right\rangle\mathtt{bdiag}(\overline{\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})})\right\|_F\\
\leq&\rho^{-1}\sqrt{n_3}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})\|_F^2\left\|\sum_{ijk}\left\langle\mathfrak{e}_{ijk},\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{Z}})\right\rangle\right\|_F\\
=&\rho^{-1}\sqrt{n_3}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})\|_F^2\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{Z}})\|_F\\
\leq&\rho^{-1}\sqrt{n_3}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})\|_F^2\|\boldsymbol{\mathcal{Z}}\|_F\\
=&\rho^{-1}\|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\mathfrak{e}_{ijk})\|_F^2\|\bar{\boldsymbol{Z}}\|_F\\
\leq&\frac{2\mu r}{nn_3\rho}\|\bar{\boldsymbol{Z}}\|_F.
\end{aligned}
$$

This implies $\left\|\sum_{ijk}\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}^2]\right\|\leq\frac{2\mu r}{nn_3\rho}$. Let $\epsilon\leq1$. By Lemma 5.1, we have

$$
\begin{aligned}
&\mathbb{P}\left[\|\rho^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}-\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}\|>\epsilon\right]\\
=&\mathbb{P}\left[\left\|\sum_{ijk}\boldsymbol{\mathcal{H}}_{ijk}\right\|>\epsilon\right]\\
=&\mathbb{P}\left[\left\|\sum_{ijk}\bar{\boldsymbol{H}}_{ijk}\right\|>\epsilon\right]\\
\leq&2nn_3\exp\left(-\frac{3}{8}\cdot\frac{\epsilon^2}{2\mu r/(nn_3\rho)}\right)\\
\leq&2(nn_3)^{1-\frac{3}{16}C_0},
\end{aligned}
$$

where the last inequality uses $\rho\geq C_0\epsilon^{-2}\mu r\log(nn_3)/(nn_3)$. Thus, $\|\rho^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}-\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}\|\leq\epsilon$ holds with high probability for some numerical constant C_0. □

5.4 Proof of Corollary 4.2

Proof. From Lemma 4.1, we have

$$\|\boldsymbol{\mathcal{P}_T} - (1-\rho)^{-1}\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_{\Omega^\perp}}\boldsymbol{\mathcal{P}_T}\| \leq \epsilon,$$

provided that $1-\rho \geq C_0\epsilon^{-2}(\mu r\log(nn_3))/n$. Note that $\boldsymbol{\mathcal{I}} = \boldsymbol{\mathcal{P}_\Omega} + \boldsymbol{\mathcal{P}_{\Omega^\perp}}$, we have

$$\|\boldsymbol{\mathcal{P}_T} - (1-\rho)^{-1}\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_{\Omega^\perp}}\boldsymbol{\mathcal{P}_T}\| = (1-\rho)^{-1}(\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T} - \rho\boldsymbol{\mathcal{P}_T}).$$

Then, by the triangular inequality

$$\|\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\| \leq \epsilon(1-\rho) + \rho\|\boldsymbol{\mathcal{P}_T}\| = \rho + \epsilon(1-\rho).$$

The proof is completed by using $\|\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\|^2 = \|\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_\Omega}\boldsymbol{\mathcal{P}_T}\|$. □

5.5 Proof of Lemma 4.3

Proof. For any tensor $\boldsymbol{\mathcal{Z}} \in \boldsymbol{T}$, we write

$$\rho^{-1}\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_\Omega}(\boldsymbol{\mathcal{Z}}) = \sum_{ijk}\rho^{-1}\delta_{ijk}z_{ijk}\boldsymbol{\mathcal{P}_T}(\mathfrak{e}_{ijk}).$$

The (a,b,c)-th entry of $\rho^{-1}\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_\Omega}(\boldsymbol{\mathcal{Z}}) - \boldsymbol{\mathcal{Z}}$ can be written as a sum of independent random variables, i.e.,

$$\begin{aligned}&\left\langle \rho^{-1}\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_\Omega}(\boldsymbol{\mathcal{Z}}) - \boldsymbol{\mathcal{Z}}, \mathfrak{e}_{abc}\right\rangle\\ =&\sum_{ijk}(\rho^{-1}\delta_{ijk}-1)z_{ijk}\left\langle\boldsymbol{\mathcal{P}_T}(\mathfrak{e}_{ijk}),\mathfrak{e}_{abc}\right\rangle\\ :=&\sum_{ijk}t_{ijk},\end{aligned}$$

where t_{ijk}'s are independent and $\mathbb{E}(t_{ijk}) = 0$. Now we bound $|t_{ijk}|$ and $|\sum_{ijk}\mathbb{E}[t_{ijk}^2]|$. First

$$\begin{aligned}&|t_{ijk}|\\ \leq&\rho^{-1}\|\boldsymbol{\mathcal{Z}}\|_\infty\|\boldsymbol{\mathcal{P}_T}(\mathfrak{e}_{ijk})\|_F\|\boldsymbol{\mathcal{P}_T}(\mathfrak{e}_{abc})\|_F\\ \leq&\frac{2\mu r}{nn_3\rho}\|\boldsymbol{\mathcal{Z}}\|_\infty.\end{aligned}$$

Second, we have

$$\begin{aligned}&\left|\sum_{ijk}\mathbb{E}[t_{ijk}^2]\right|\\ \leq&\rho^{-1}\|\boldsymbol{\mathcal{Z}}\|_\infty^2\sum_{ijk}\left\langle\boldsymbol{\mathcal{P}_T}(\mathfrak{e}_{ijk}),\mathfrak{e}_{abc}\right\rangle^2\\ =&\rho^{-1}\|\boldsymbol{\mathcal{Z}}\|_\infty^2\sum_{ijk}\left\langle\mathfrak{e}_{ijk},\boldsymbol{\mathcal{P}_T}(\mathfrak{e}_{abc})\right\rangle^2\\ =&\rho^{-1}\|\boldsymbol{\mathcal{Z}}\|_\infty^2\|\boldsymbol{\mathcal{P}_T}(\mathfrak{e}_{abc})\|_F^2\\ \leq&\frac{2\mu r}{nn_3\rho}\|\boldsymbol{\mathcal{Z}}\|_\infty^2.\end{aligned}$$

Let $\epsilon \leq 1$. By Lemma 5.1, we have

$$\begin{aligned}&\mathbb{P}\left[|[\rho^{-1}\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_\Omega}(\boldsymbol{\mathcal{Z}}) - \boldsymbol{\mathcal{Z}}]_{abc}| > \epsilon\|\boldsymbol{\mathcal{Z}}\|_\infty\right]\\ =&\mathbb{P}\left[\left|\sum_{ijk}t_{ijk}\right| > \epsilon\|\boldsymbol{\mathcal{Z}}\|_\infty\right]\\ \leq&2\exp\left(-\frac{3}{8}\cdot\frac{\epsilon^2\|\boldsymbol{\mathcal{Z}}\|_\infty^2}{2\mu r\|\boldsymbol{\mathcal{Z}}\|_\infty^2/(nn_3\rho)}\right)\\ \leq&2(nn_3)^{-\frac{3}{16}C_0},\end{aligned}$$

where the last inequality uses $\rho \geq C_0\epsilon^{-2}\mu r\log(nn_3)/(nn_3)$. Thus, $\|\rho^{-1}\boldsymbol{\mathcal{P}_T}\boldsymbol{\mathcal{P}_\Omega}(\boldsymbol{\mathcal{Z}}) - \boldsymbol{\mathcal{Z}}\|_\infty \leq \epsilon\|\boldsymbol{\mathcal{Z}}\|_\infty$ holds with high probability for some numerical constant C_0. □

5.6 Proof of Lemma 4.4

Proof. Denote the tensor $\boldsymbol{\mathcal{H}}_{ijk} = \left(1-\rho^{-1}\delta_{ijk}\right)z_{ijk}\mathfrak{e}_{ijk}$. Then we have

$$(\boldsymbol{\mathcal{I}} - \rho^{-1}\boldsymbol{\mathcal{P}_\Omega})\boldsymbol{\mathcal{Z}} = \sum_{ijk}\boldsymbol{\mathcal{H}}_{ijk}.$$

Note that δ_{ijk}'s are independent random scalars. Thus, $\boldsymbol{\mathcal{H}}_{ijk}$'s are independent random tensors and $\bar{\boldsymbol{H}}_{ijk}$'s are independent random matrices. Observe that $\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}] = \boldsymbol{0}$ and $\|\bar{\boldsymbol{H}}_{ijk}\| \leq \rho^{-1}\|\boldsymbol{\mathcal{Z}}\|_\infty$. We have

$$\begin{aligned}&\left\|\sum_{ijk}\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}^*\bar{\boldsymbol{H}}_{ijk}]\right\|\\ =&\left\|\sum_{ijk}\mathbb{E}[\boldsymbol{\mathcal{H}}_{ijk}^* * \boldsymbol{\mathcal{H}}_{ijk}]\right\|\\ =&\left\|\sum_{ijk}\mathbb{E}[(1-\rho^{-1}\delta_{ijk})^2]z_{ijk}^2(\mathring{\mathfrak{e}}_j * \mathring{\mathfrak{e}}_j^*)\right\|\\ =&\left\|\frac{1-\rho}{\rho}\sum_{ijk}z_{ijk}^2(\mathring{\mathfrak{e}}_j * \mathring{\mathfrak{e}}_j^*)\right\|\\ \leq&\frac{nn_3}{\rho}\|\boldsymbol{\mathcal{Z}}\|_\infty^2.\end{aligned}$$

A similar calculation yields $\left\|\sum_{ijk}\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}^*\bar{\boldsymbol{H}}_{ijk}]\right\| \leq \rho^{-1}nn_3\|\boldsymbol{\mathcal{Z}}\|_\infty^2$. Let $t = \sqrt{C_0 nn_3\log(nn_3)/\rho}\|\boldsymbol{\mathcal{Z}}\|_\infty$. When

$\rho \geq C_0 \log(nn_3)/(nn_3)$, we apply Lemma 5.1 and obtain

$$
\begin{aligned}
&\mathbb{P}\left[\|(\boldsymbol{\mathcal{I}} - \rho^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}})\boldsymbol{\mathcal{Z}}\| > t\right] \\
=&\mathbb{P}\left[\left\|\sum_{ijk} \boldsymbol{\mathcal{H}}_{ijk}\right\| > t\right] \\
=&\mathbb{P}\left[\left\|\sum_{ijk} \bar{\boldsymbol{H}}_{ijk}\right\| > t\right] \\
\leq& 2nn_3 \exp\left(-\frac{3}{8} \cdot \frac{C_0 nn_3 \log(nn_3)\|\boldsymbol{\mathcal{Z}}\|_\infty^2/\rho}{nn_3\|\boldsymbol{\mathcal{Z}}\|_\infty^2/\rho}\right) \\
\leq& 2(nn_3)^{1-\frac{3}{8}C_0}.
\end{aligned}
$$

Thus, $\|(\boldsymbol{\mathcal{I}} - \rho^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}})\boldsymbol{\mathcal{Z}}\| > t$ holds with high probability for some numerical constant C_0. □

References

[1] E. J. Candès, X. D. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3), 2011. 1

[2] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012. 6

[3] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010. 4, 5

[4] G. A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170:33–45, 1992. 2