Journal of Machine Learning Research 24 (2023) 1-84 Submitted 11/22; Revised 10/23; Published 11/23

Implicit Regularization and Entrywise Convergence of Riemannian Optimization for Low Tucker-Rank Tensor Completion

Haifeng Wang* 20110980008@FUDAN.EDU.CN
School of Data Science
Fudan University
Shanghai, China

China Mobile (Zhejiang) Research & Innovation Institute
Hangzhou, China

Jinchi Chen* JCCHEN.PHYS@GMAIL.COM
School of Mathematics
East China University of Science and Technology
Shanghai, China

Ke Wei KEWEI@FUDAN.EDU.CN
School of Data Science
Fudan University
Shanghai, China

Editor: Daniel Hsu

Abstract

This paper is concerned with the low Tucker-rank tensor completion problem, which is about reconstructing a tensor $\mathcal{T} \in \mathbb{R}^{n\times n\times n}$ of low multilinear rank from partially observed entries. Riemannian optimization algorithms are a class of efficient methods for this problem, but the theoretical convergence analysis is still lacking. In this manuscript, we establish the entrywise convergence of the vanilla Riemannian gradient method for low Tucker-rank tensor completion under the nearly optimal sampling complexity $O(n^{3/2})$. Meanwhile, the implicit regularization phenomenon of the algorithm has also been revealed. As far as we know, this is the first work that has shown the entrywise convergence and implicit regularization property of a non-convex method for low Tucker-rank tensor completion. The analysis relies on the leave-one-out technique, and some of the technical results developed in the paper might be of broader interest in investigating the properties of other non-convex methods for this problem.

Keywords: low rank tensor completion, Tucker decomposition, Riemannian gradient, entrywise convergence, implicit regularization, leave-one-out

1. Introduction

Tensors are multidimensional arrays which are ubiquitous in data analysis, including but not limited to topic modeling (Anandkumar et al., 2015), community detection (Anandkumar

*. Haifeng Wang and Jinchi Chen contribute equally to this work.

et al., 2013), computer version (Liu et al., 2012), collaborative filtering (Karatzoglou et al., 2010), and signal processing (Cichocki et al., 2015). In this paper, we consider the tensor completion problem which is about reconstructing a tensor from a few observed entries. Without any additional assumptions, tensor completion is an ill-posed problem which does not even have a unique solution. Therefore computationally efficient solution of this problem is typically based on certain intrinsic low dimensional structures of tensors, a notable example of which is low rank. Compared with matrix, tensor has more complex rank notions up to different tensor decompositions such as CANDECOMP/PARAFAC (CP) decomposition (Hitchcock, 1927), Tucker decomposition (Tucker, 1966), tensor train (TT) decomposition (Oseledets, 2011), and t-SVD decomposition (Zhang and Aeron, 2016). In this manuscript, we focus on the Tucker decomposition. Then the low-rank tensor completion problem can be formulated as follows:

$$\min_{\mathcal{X}\in\mathbb{R}^{n\times n\times n}} \frac{1}{2p}\left\|\mathcal{P}_\Omega\left(\mathcal{X}\right)-\mathcal{P}_\Omega\left(\mathcal{T}\right)\right\|_{\mathsf{F}}^2, \quad \text{s.t.} \quad \operatorname{rank}\left(\mathcal{X}\right)=\boldsymbol{r} \tag{1}$$

where $\mathcal{T}\in\mathbb{R}^{n\times n\times n}$ is the target tensor to be recovered, $\operatorname{rank}\left(\mathcal{X}\right)$ is the Tucker rank of $\mathcal{X}$ which will be specified later, Ω is a subset of indices for the observed entries, $p=|\Omega|/n^3$ is the sampling rate, $\mathcal{P}_\Omega$ is the element-wise sampling operator, and $\|\cdot\|_{\mathsf{F}}$ denotes the Frobenius norm (e.g., $\|\mathcal{X}\|_{\mathsf{F}}^2=\sum_{i_1,i_2,i_3}\mathcal{X}_{i_1,i_2,i_3}^2$).

1.1 Main contributions

For the low rank tensor completion problem under Tucker decomposition, many methods have been developed (Gandy et al., 2011; Huang et al., 2015; Han et al., 2020; Kressner et al., 2014; Liu et al., 2012; Luo and Zhang, 2021; Mu et al., 2014; Xia and Yuan, 2019; Rauhut et al., 2017; Tong et al., 2021). Among them, Riemannian optimization algorithms are a class of efficient methods. Despite the computational efficiency of Riemannian optimization, there still lacks theoretical analysis for them. In this manuscript, we fill this gap by providing an entrywise convergence of the vanilla Riemannian gradient method (RGM) for low rank tensor completion.

RGM for tensor completion is an extension of the method for matrix completion (Vandereycken, 2013; Wei et al., 2016, 2020). The iteration of RGM is given by

$$\mathcal{X}^{t+1}=\operatorname{HOSVD}\left(\mathcal{X}^t-p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_\Omega\left(\mathcal{X}^t-\mathcal{T}\right)\right),$$

where the retraction HOSVD and the projection $\mathcal{P}_{T_{\mathcal{X}^t}}$ are specified in Section 2.1. Assume the target tensor $\mathcal{T}$ is incoherent, i.e., there exists a constant $\mu>0$ such that

$$\|\boldsymbol{U}_i\|_{2,\infty}\le\sqrt{\frac{\mu r}{n}}, \quad i=1,2,3,$$

where $\boldsymbol{U}_i\in\mathbb{R}^{n\times r}$ $(i=1,2,3)$ are the factor matrices of the tensor $\mathcal{T}$, and $\|\boldsymbol{U}_i\|_{2,\infty}:=\max_{j\in[n]}\left\|\left[\boldsymbol{U}_i\right]_{j,:}\right\|_2$ is the $\ell_{2,\infty}$ norm of $\boldsymbol{U}_i$. Let $\mathcal{X}^t\in\mathbb{R}^{n\times n\times n}$ with factor matrices $\boldsymbol{X}_i^t\in\mathbb{R}^{n\times r}$ $(i=1,2,3)$ denote the iterate generated by the RGM in the t-th iteration. We have the following main result.

Theorem 1 *Assume each entry of $\mathcal{T}$ is observed independently with probability p. Let the condition number of $\mathcal{T}$ be κ. If $p \geq \frac{O(r^6\kappa^8\log^3 n)}{n^{3/2}}$, the iterates of the Riemannian gradient method (Algorithm 2) satisfy the following properties with high probability:*

1. *Incoherence of the iterates:*
$$\left\|\boldsymbol{X}_i^t\right\|_{2,\infty} \leq 2\kappa\sqrt{\frac{\mu r}{n}}, \quad t \geq 1. \tag{2}$$

2. *The iterates $\mathcal{X}^t$ converges linearly to $\mathcal{T}$ in terms of the infinity norm:*
$$\left\|\mathcal{X}^t - \mathcal{T}\right\|_\infty \leq \left(\frac{1}{2}\right)^t \sigma_{\max}\left(\mathcal{T}\right), \quad t \geq 1, \tag{3}$$
where $\sigma_{\max}\left(\mathcal{T}\right)$ denotes the upper bound on the largest singular values of the matricization of $\mathcal{T}$ along every mode (see Equation 5), and $\left\|\mathcal{X}^t - \mathcal{T}\right\|_\infty := \max_{i_1,i_2,i_3}\left|(\mathcal{X}^t - \mathcal{T})_{i_1,i_2,i_3}\right|$ denotes the infinity norm of $\mathcal{X}^t - \mathcal{T}$. In particular, for the first logarithm number of iterations, a stronger entrywise convergence of RGM can be established (see Theorem 6):
$$\left\|\mathcal{X}^t - \mathcal{T}\right\|_\infty \leq \left(\frac{1}{2}\right)^t \frac{1}{n^{3/2}}\sigma_{\max}\left(\mathcal{T}\right), \quad \textit{for } 1 \leq t \leq t_0, \ \textit{where } t_0 = O(\log_2 n).$$

Remark 2 *Though the results are established for the Bernoulli model, they equally apply to the uniform sampling model without replacement as the two models are closely related to each other. The Bernoulli model is adopted for considerably simpler argument, as is done in related literature (Recht, 2011; Yuan and Zhang, 2016; Xia and Yuan, 2019).*

Remark 3 *As a byproduct, the convergence of the vanilla RGM for low rank matrix completion follows immediately from Theorem 1. This is also a new result for the matrix case since previous results either require a stronger initialization procedure (Wei et al., 2020) or require an additional projection in the algorithm (Cai et al., 2021c). The analyses in Wei et al., 2020; Cai et al., 2021c are all based on the Frobenius norm error metric, which cannot fully exploit the incoherence property of each iterate and thus require sample splitting for initialization (so that the initialization can be unreasonably close to the ground truth) or explicit projection onto the incoherence region in the algorithm. However, these additional steps are empirically not necessary. In contrast, the analysis in this paper is based on the $\ell_{2,\infty}$ norm and infinity norm which enables us to analyse the incoherence property of each iterate more carefully and establish the convergence of the algorithm that does not have those empirically redundant steps.*

Inequality (2) shows that iterates of the RGM remain incoherent even in the absence of explicit regularization which is known as the *implicit regularization* phenomenon. To the best of our knowledge, this is the first work that has shown the stronger entrywise convergence and the implicit regularization phenomenon of a non-convex method for low Tucker-rank tensor completion under the nearly optimal sampling complexity $O(n^{3/2})$. Table 1 summarizes the theoretical recovery guarantees of different nonconvex algorithms for both the Gaussian measurement model and the entrywise measurement model.

Table 1: Theories of different nonconvex methods for low Tucker-rank tensor completion.

Algorithms	Sampling complexity	Error metric	Sampling scheme
Projected GD (Chen et al., 2019)	$n^2 r$	Frobenius	Gaussian
Regularized GD (Han et al., 2020)	$n^{3/2} r \kappa^4$	Frobenius	Gaussian
RGM (Cai et al., 2020)	$n^{3/2} r^2 \kappa^2$	Frobenius	Gaussian
Riemannian Gauss-Newton (Luo and Zhang, 2021)	$n^{3/2} r^{3/2} \kappa^4$	Frobenius	Gaussian
ScaledGD (Tong et al., 2021)	$n^{3/2} r \kappa^2$	Frobenius	Gaussian
Grassmannian GD (Xia and Yuan, 2019)	$n^{3/2} r^{7/2} \kappa^4 \log^{7/2} n$	Frobenius	Entrywise
ScaledGD (Tong et al., 2021)	$n^{3/2} r^2 \kappa \left(\sqrt{r} \vee \kappa^2\right) \log^3 n$	Frobenius	Entrywise
RGM [**this paper**]	$n^{3/2} r^6 \kappa^8 \log^3 n$	**Infinity**	Entrywise

The proof of Theorem 1 relies on the leave-one-out technique which has been widely used in analyzing various high dimensional data processing methods. Our work is mostly inspired by Cai et al., 2021b for low rank tensor completion problem under the CP decomposition and by Ding and Chen, 2020 for low rank matrix completion, but the technical details are substantially different. On the one hand, CP decomposition and Tucker decomposition are essentially two different decompositions for tensors and a gradient descent algorithm is analysed in Cai et al., 2021b. Even though a few results for the initialization step therein can be used in our proof, the proof details for the iteration procedure are significantly different. Even in the initialization step, we find that the structure of Tucker decomposition can be used to simplify the proof of a related result (Lemma 16). On the other hand, tensors are more complex than matrices which means that, compared with the analysis of iterative hard thresholding (IHT) for low rank matrix completion, the analysis of RGM for tensor completion are much more complicated. Moreover, differing from IHT, there is one key different component in RGM, namely the orthogonal projection $\mathcal{P}_{T_{\mathcal{X}^t}}$. Indeed, we need to carefully leverage this projection to obtain the nearly optimal sampling complexity $O(n^{3/2})$. Without this projection step, the near-optimal sampling complexity can not be achieved for the low rank tensor completion problem. This is basically because in the tensor case we need to transform the perturbation tensor into an $n \times n^2$ matrix via matricization in order to bound it and exploiting the low dimensional structure is essential for us to reduce the dependence on n from n^2 to $n^{3/2}$. In contrast, for the low rank matrix completion problem, the convergence result can be established with or without this projection step since it does not involve the transformation from an tensor to an $n \times n^2$ matrix.

1.2 Organization and Notation

This paper is organized as follows. Section 2 begins with some tensor preliminaries and then introduces the Riemannian gradient method for low Tucker-rank tensor completion together with the main theoretical results. The proof strategies of the main results are outlined in Sections 3 and 4, while the details are provided in the appendix. In Section 5, we conclude this paper with a few future directions.

Throughout this paper, tensors are denoted by capital calligraphic letters, matrices are denoted by bold capital letters, and vectors are denoted by bold lower case letters. In particular, we will reserve $\mathcal{T}$ for the ground truth tensor to be recovered, which is an $n \times n \times n$ tensor of multilinear rank $\boldsymbol{r} = (r_1, r_2, r_3)$. For ease of exposition, we assume $r_1 = r_2 = r_3 = r$ for the ground truth tensor. For each integer d, define the set $[d] = \{1, 2, \cdots, d\}$. We denote by $\boldsymbol{e}_i$ the i-th canonical basis vector, by $\boldsymbol{I}$ the identity matrix with suitable size, and by $\mathcal{J}$ the all-one tensor (i.e., all the entries of $\mathcal{J}$ are 1). For any matrix $\boldsymbol{M}$, we use $[\boldsymbol{M}]_{i,:}$, $[\boldsymbol{M}]_{:,j}$, $[\boldsymbol{M}]_{i,j}$ to represent the i-th row, the j-th column, the (i,j)-th element of $\boldsymbol{M}$, respectively. The Frobenius norm, spectral norm and nuclear norm of a matrix $\boldsymbol{M}$ are denoted by $\|\boldsymbol{M}\|_{\mathsf{F}}$, $\|\boldsymbol{M}\|$, $\|\boldsymbol{M}\|_*$, respectively. In addition, the $\ell_{2,\infty}$ norm of a matrix $\boldsymbol{M}$ is defined as $\|\boldsymbol{M}\|_{2,\infty} := \max_{i\in[n]} \left\| [\boldsymbol{M}]_{i,:} \right\|_2$. For a tensor $\mathcal{X}$, its infinity norm is defined as $\|\mathcal{X}\|_\infty := \max_{i_1,i_2,i_3} |\mathcal{X}_{i_1,i_2,i_3}|$. We will use $C, C_0, C_1, \cdots$ to denote absolute positive constants, whose values may vary from line to line. Lastly, we use the terminology "with high probability" to denote the event happens with probability at least $1 - C_1 n^{-C_2}$ for some constants $C_1, C_2 > 0$ and C_2 sufficiently large.

2. Algorithm and Main Result

2.1 Preliminaries

We begin this section with some preliminaries of tensors; for a more detailed exposition, see Kolda and Bader (2009). For conciseness, we restrict our discussion to $n \times n \times n$ three-way tensors.

Tensor matricization. Matricization, also known as unfolding, transforms a tensor into a matrix along different modes. Given a tensor $\mathcal{X} \in \mathbb{R}^{n\times n\times n}$, the matricization operators are defined as

$$\begin{aligned}
\mathcal{M}_1(\mathcal{X}) \in \mathbb{R}^{n\times n^2} &: [\mathcal{M}_1(\mathcal{X})]_{i_1, i_2+n(i_3-1)} = \mathcal{X}_{i_1,i_2,i_3},\\
\mathcal{M}_2(\mathcal{X}) \in \mathbb{R}^{n\times n^2} &: [\mathcal{M}_2(\mathcal{X})]_{i_2, i_1+n(i_3-1)} = \mathcal{X}_{i_1,i_2,i_3},\\
\mathcal{M}_3(\mathcal{X}) \in \mathbb{R}^{n\times n^2} &: [\mathcal{M}_3(\mathcal{X})]_{i_3, i_1+n(i_2-1)} = \mathcal{X}_{i_1,i_2,i_3}.
\end{aligned}$$

Mode-d tensor multiplication. The mode-1 product of a tensor $\mathcal{X} \in \mathbb{R}^{n\times n\times n}$ with a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, denoted $\mathcal{X} \times_1 \boldsymbol{A}$, gives a tensor of size $m \times n \times n$. Elementwise, we have

$$(\mathcal{X} \times_1 \boldsymbol{A})_{j_1,i_2,i_3} = \sum_{i_1=1}^{n} \mathcal{X}_{i_1,i_2,i_3} [\boldsymbol{A}]_{j_1,i_1} ,$$

and $\times_2$ and $\times_3$ are similarly defined. A few facts regarding mode-d tensor multiplication are in order:

$$\mathcal{X} \times_i \boldsymbol{A} \times_j \boldsymbol{B} = \mathcal{X} \times_j \boldsymbol{B} \times_i \boldsymbol{A} \ (i \neq j) \quad \text{and} \quad \mathcal{X} \times_i \boldsymbol{A} \times_i \boldsymbol{B} = \mathcal{X} \times_i (\boldsymbol{B}\boldsymbol{A}).$$

Tensor norms. The inner product between two tensors is defined as

$$\langle \mathcal{X}, \mathcal{Z} \rangle := \sum_{i_1, i_2, i_3} \mathcal{X}_{i_1,i_2,i_3} \cdot \mathcal{Z}_{i_1,i_2,i_3}.$$

The Frobenius norm and the spectral norm of a tensor are defined as

$$\|\mathcal{X}\|_{\mathsf{F}} := \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle} \quad \text{and} \quad \|\mathcal{X}\| := \sup_{\boldsymbol{x}_i \in \mathbb{R}^n : \|\boldsymbol{x}_i\|_2 = 1} \langle \mathcal{X}, \boldsymbol{x}_1 \circ \boldsymbol{x}_2 \circ \boldsymbol{x}_3 \rangle,$$

where the element of $\boldsymbol{x}_1 \circ \boldsymbol{x}_2 \circ \boldsymbol{x}_3 \in \mathbb{R}^{n \times n \times n}$ is defined by

$$[\boldsymbol{x}_1 \circ \boldsymbol{x}_2 \circ \boldsymbol{x}_3]_{i_1,i_2,i_3} := [\boldsymbol{x}_1]_{i_1} \cdot [\boldsymbol{x}_2]_{i_2} \cdot [\boldsymbol{x}_3]_{i_3}. \tag{4}$$

The following basic relations regarding the spectral norm, which follow immediately from the definition, will be very useful: for any $i \in [3]$,

$$\|\mathcal{X}\| \leq \|\mathcal{M}_i(\mathcal{X})\| \quad \text{and} \quad \|\mathcal{X} \times_i \boldsymbol{X}\| \leq \|\boldsymbol{X}\| \cdot \|\mathcal{X}\|.$$

Similar to the matrix case, the nuclear norm is the dual of spectral norm:

$$\|\mathcal{X}\|_* := \sup_{\mathcal{Z} \in \mathbb{R}^{n \times n \times n}, \|\mathcal{Z}\| \leq 1} \langle \mathcal{X}, \mathcal{Z} \rangle.$$

Recall that the condition number for a matrix $\boldsymbol{A}$ is given by $\kappa(\boldsymbol{A}) = \sigma_{\max}(\boldsymbol{A}) / \sigma_{\min}(\boldsymbol{A})$ where $\sigma_{\max}$ and $\sigma_{\min}$ are the largest and smallest nonzero singular values of $\boldsymbol{A}$ respectively. This concept can be naturally generalized to tensors,

$$\kappa(\mathcal{X}) := \frac{\sigma_{\max}(\mathcal{X})}{\sigma_{\min}(\mathcal{X})},$$

where $\sigma_{\max}(\mathcal{X})$ and $\sigma_{\min}(\mathcal{X})$ are defined as

$$\sigma_{\max}(\mathcal{X}) := \max\{\sigma_{\max}(\mathcal{M}_1(\mathcal{X})), \sigma_{\max}(\mathcal{M}_2(\mathcal{X})), \sigma_{\max}(\mathcal{M}_3(\mathcal{X}))\} \tag{5}$$

$$\sigma_{\min}(\mathcal{X}) := \min\{\sigma_{\min}(\mathcal{M}_1(\mathcal{X})), \sigma_{\min}(\mathcal{M}_2(\mathcal{X})), \sigma_{\min}(\mathcal{M}_3(\mathcal{X}))\}. \tag{6}$$

Tucker decomposition and HOSVD. The Tucker decomposition is a higher-order generalization of singular value decomposition (SVD), which has the form

$$\mathcal{X} = \mathcal{G} \times_1 \boldsymbol{X}_1 \times_2 \boldsymbol{X}_2 \times_3 \boldsymbol{X}_3 = \mathcal{G} \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i, \tag{7}$$

where $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ is referred to as the core tensor, and $\boldsymbol{X}_i \in \mathbb{R}^{n \times r_i}$ for $i = 1, 2, 3$ are the factor matrices. Because $\mathcal{G}$ is usually unstructured, we can always write the Tucker decomposition of $\mathcal{X}$ into the form where $\boldsymbol{X}_i$ are orthonormal matrices. Given the Tucker decomposition of $\mathcal{X}$, its matricizations are given by

$$\begin{aligned}\mathcal{M}_1(\mathcal{X}) &= \boldsymbol{X}_1\mathcal{M}_1(\mathcal{G})(\boldsymbol{X}_3\otimes\boldsymbol{X}_2)^{\mathsf{T}},\\ \mathcal{M}_2(\mathcal{X}) &= \boldsymbol{X}_2\mathcal{M}_2(\mathcal{G})(\boldsymbol{X}_3\otimes\boldsymbol{X}_1)^{\mathsf{T}},\\ \mathcal{M}_3(\mathcal{X}) &= \boldsymbol{X}_3\mathcal{M}_3(\mathcal{G})(\boldsymbol{X}_2\otimes\boldsymbol{X}_1)^{\mathsf{T}},\end{aligned}$$

where $\otimes$ denotes the Kronecker product of matrices.

A tensor $\mathcal{X}\in\mathbb{R}^{n\times n\times n}$ is said to be of multilinear rank $\boldsymbol{r}=(r_1,r_2,r_3)$ if $\operatorname{rank}(\mathcal{M}_i(\mathcal{X}))=r_i$ for $i=1,2,3$. It is evident that $1\le r_i\le n$ for $i=1,2,3$. It is well known that SVD can be used to find the best low rank approximation of a matrix. In contrast, computing the best low rank approximation of a tensor is an NP hard (Hillar and Lim, 2013) problem. That being said, there exists a higher order analogue of SVD, known as Higher Order Singular Value Decomposition (HOSVD) (De Lathauwer et al., 2000), which is able to return a quasi-optimal approximation; see Algorithm 1. HOSVD first estimates the principle factor matrices of each mode by an SVD tunctation of the corresponding matricization, and then formulates the core tensor by multiplying $\mathcal{X}$ by the transpose of the factor matrix along each mode. Denoting by $\mathcal{H}_{\boldsymbol{r}}(\mathcal{X})$ the output of HOSVD, there holds (De Lathauwer et al., 2000)

$$\|\mathcal{X}-\mathcal{H}_{\boldsymbol{r}}(\mathcal{X})\|_{\mathsf{F}}\le\sqrt{3}\inf_{\operatorname{rank}(\mathcal{M}_i(\mathcal{Z}))\le r_i}\|\mathcal{X}-\mathcal{Z}\|_{\mathsf{F}}.$$

Note that when $\mathcal{X}$ is already of multilinear rank $\boldsymbol{r}$, HOSVD returns the exact Tucker decomposition of $\mathcal{X}$.

Algorithm 1 HOSVD

1: Input: Tensor $\mathcal{X}\in\mathbb{R}^{n\times n\times n}$, multilinear rank $\boldsymbol{r}=(r_1,r_2,r_3)$.
2: **for** $i=1,2,3$ **do**
3: $\quad\boldsymbol{X}_i=\mathrm{SVD}_{r_i}(\mathcal{M}_i(\mathcal{X}))$.
4: **end for**
5: $\mathcal{G}=\mathcal{X}\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i^{\mathsf{T}}$.
6: Output: $\mathcal{G}\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i$.

Tensor manifold. A collection of tensors with multilinear rank $\boldsymbol{r}=(r_1,r_2,r_3)$ forms a smooth embedded submanifold of $\mathbb{R}^{n\times n\times n}$ (Koch and Lubich, 2010), denoted $\mathbb{M}_{\boldsymbol{r}}$, i.e.,

$$\mathbb{M}_{\boldsymbol{r}}=\left\{\mathcal{X}\in\mathbb{R}^{n\times n\times n}\ \mid\ \operatorname{rank}(\mathcal{X})=\boldsymbol{r}\right\}.$$

Let the Tucker decomposition of $\mathcal{X}$ be $\mathcal{X}=\mathcal{G}\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i$, where $\boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{X}_i=\boldsymbol{I}\in\mathbb{R}^{r_i\times r_i}$ and $\mathcal{G}\in\mathbb{R}^{r_1\times r_2\times r_3}$ has full multilinear rank. The tangent space of $\mathbb{M}_{\boldsymbol{r}}$ at $\mathcal{X}$ is given by Koch and Lubich, 2010:

$$T_{\mathcal{X}}=\left\{\mathcal{C}\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i+\sum_{i=1}^{3}\mathcal{G}\times_i\boldsymbol{W}_i\underset{j\neq i}{\times}\boldsymbol{X}_j\ \mid\ \mathcal{C}\in\mathbb{R}^{r_1\times r_2\times r_3},\boldsymbol{W}_i\in\mathbb{R}^{n\times r_i},\boldsymbol{W}_i^{\mathsf{T}}\boldsymbol{X}_i=\boldsymbol{0},i=1,2,3\right\}.$$

Given a tensor $\mathcal{Z}$, the orthogonal projection of $\mathcal{Z}$ onto $T_{\mathcal{X}}$, denoted $\mathcal{P}_{T_{\mathcal{X}}}(\mathcal{Z})$, has the form $\mathcal{P}_{T_{\mathcal{X}}}(\mathcal{Z}) = \mathcal{C} \times_{i=1}^{3} \boldsymbol{X}_i + \sum_{i=1}^{3} \mathcal{G} \times_i \boldsymbol{W}_i \times_{j\neq i} \boldsymbol{X}_j$, where $\mathcal{C}$ and $\boldsymbol{W}_i$ are to be determined. Since the summands are orthogonal to each other, $\mathcal{C}$ and $\boldsymbol{W}_i$ can be obtained independently by solving the least squares problems, yielding the expression (Koch and Lubich, 2010; Kressner et al., 2014; Cai et al., 2020)

$$\mathcal{P}_{T_{\mathcal{X}}}(\mathcal{Z}) = \mathcal{Z} \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i \boldsymbol{X}_i^{\mathsf{T}} + \sum_{i=1}^{3} \mathcal{G} \times_i \boldsymbol{W}_i \mathop{\times}_{j\neq i} \boldsymbol{X}_j, \tag{8}$$

where $\boldsymbol{W}_i$ is defined as

$$\boldsymbol{W}_i = \left(\boldsymbol{I} - \boldsymbol{X}_i \boldsymbol{X}_i^{\mathsf{T}}\right) \mathcal{M}_i \left(\mathcal{Z} \mathop{\times}_{j\neq i} \boldsymbol{X}_j^{\mathsf{T}}\right) \mathcal{M}_i^{\dagger}(\mathcal{G}).$$

Here $\mathcal{M}_i^{\dagger}(\mathcal{G})$ denotes the Moore-Penrose pseudoinverse of $\mathcal{M}_i(\mathcal{G})$ (Golub and Van Loan, 1996), which obeys that $\mathcal{M}_i(\mathcal{G})\mathcal{M}_i^{\dagger}(\mathcal{G}) = \boldsymbol{I}$ and $\mathcal{M}_i^{\dagger}(\mathcal{G})\mathcal{M}_i(\mathcal{G})$ is an orthogonal projector.

2.2 Riemannian Gradient Method

Algorithm 2 Riemannian Gradient Method (RGM)

1: Input: Initialization $\mathcal{X}^1$ generated via Algorithm 3, multilinear rank $\boldsymbol{r} = (r, r, r)$, parameter p.
2: **for** $t = 1, \cdots$ **do**
3:     **for** $i = 1, 2, 3$ **do**
4:         $\boldsymbol{X}_i^{t+1} = \mathrm{SVD}_r\left(\mathcal{M}_i\left(\mathcal{X}^t - p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_{\Omega}\left(\mathcal{X}^t - \mathcal{T}\right)\right)\right)$,
5:     **end for**
6:     $\mathcal{G}^{t+1} = \left(\mathcal{X}^t - p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_{\Omega}\left(\mathcal{X}^t - \mathcal{T}\right)\right) \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i^{t+1\mathsf{T}}$.
7:     $\mathcal{X}^{t+1} = \mathcal{G}^{t+1} \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i^{t+1}$.
8: **end for**

The Riemannian gradient method (RGM) for solving (1) is presented in Algorithm 2. Let $\mathcal{X}^t$ be the current estimator, and $T_{\mathcal{X}^t}$ be the tangent space of the rank $\boldsymbol{r}$ tensor manifold at $\mathcal{X}^t$. RGM first updates $\mathcal{X}^t$ along $\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_{\Omega}\left(\mathcal{X}^t - \mathcal{T}\right)$, the gradient descent direction projected onto the tangent space $T_{\mathcal{X}^t}$, using the fixed step size 1. Then the new estimator $\mathcal{X}^{t+1}$ is obtained by projecting the update to the set of rank $\boldsymbol{r}$ tensors via HOSVD. Note that in description of Algorithm 2, we have included the details of HOSVD.

2.2.1 Initialization by spectral method with diagonal deletion

Spectral method is a widely used initialization method in matrix and tensor computation problems (Chen et al., 2020b), which typically involves the estimation of certain principal subspace from the data matrix or equivalently the Gram matrix corresponding to the data matrix. In some statistical settings, directly using the data matrix may lead to a biased estimator. Though this is usually not a problem when matrices are nearly square (for

Algorithm 3 Initialization via spectral method with diagonal deletion

1: Input: $\mathcal{P}_{\Omega}(\mathcal{T}) \in \mathbb{R}^{n\times n\times n}$, multilinear rank $\boldsymbol{r}=(r,r,r)$, parameter p.
2: **for** $i=1,2,3$ **do**
3: Let $\boldsymbol{X}_i^1\boldsymbol{\Sigma}_i^1\boldsymbol{X}_i^{1\top}$ be the top-r eigenvalue decomposition of $\mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top}\right)$.
4: **end for**
5: $\mathcal{G}^1 = p^{-1}\mathcal{P}_{\Omega}(\mathcal{T}) \times_{i=1}^{3} \boldsymbol{X}_i^{1\top}$.
6: Output: $\mathcal{X}^1 = \mathcal{G}^1 \times_{i=1}^{3} \boldsymbol{X}_i^1$.

example in typical matrix recovery problems (Jain et al., 2013; Keshavan et al., 2010; Ma et al., 2020)), it does lead to sub-optimal performance for highly unbalanced matrices unless the number of observations is unnecessarily large. This is the case for the tensor completion problem since for an $n\times n\times n$ tensor, we need to consider its matricization in the analysis which is an $n\times n^2$ matrix and hence highly unbalanced.

Specifically for the problem considered in this paper, the sample Gram matrix is given by $\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top}$ where $\widehat{\boldsymbol{T}}_i := p^{-1}\mathcal{M}_i(\mathcal{P}_{\Omega}(\mathcal{T})) \in \mathbb{R}^{n\times n^2}$ is the scaled observation matrix. It is clear that how close the top-r eigenvectors of the Gram matrix are to the target eigenvectors is determined by $\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top} - \boldsymbol{T}_i\boldsymbol{T}_i^{\top}$. Such error term can be decomposed into bias part and unbiased parts as follows:

$$\begin{aligned}\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top} - \boldsymbol{T}_i\boldsymbol{T}_i^{\top} &= \widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top} - \mathbb{E}\left[\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top}\right] + \mathbb{E}\left[\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top}\right] - \boldsymbol{T}_i\boldsymbol{T}_i^{\top}\\ &= \underbrace{\mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top} - \boldsymbol{T}_i\boldsymbol{T}_i^{\top}\right)}_{\text{unbiased}} + \underbrace{\mathcal{P}_{\text{diag}}\left(\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top} - p^{-1}\boldsymbol{T}_i\boldsymbol{T}_i^{\top}\right)}_{\text{unbiased}} + \underbrace{\left(p^{-1}-1\right)\mathcal{P}_{\text{diag}}\left(\boldsymbol{T}_i\boldsymbol{T}_i^{\top}\right)}_{\text{bias}},\end{aligned}$$

where $\mathcal{P}_{\text{diag}}(\boldsymbol{M})$ sets the non-diagonal elements of $\boldsymbol{M}$ to zeros and $\mathcal{P}_{\text{off-diag}}(\boldsymbol{M}) = \boldsymbol{M} - \mathcal{P}_{\text{diag}}(\boldsymbol{M})$. It can be shown that the bias term and the unbiased diagonal part lead to an unnecessarily large number of samples to ensure a reliable estimator (Florescu and Perkins, 2016; Zhang et al., 2018). To deal with this difficulty, we adopt the diagonal deletion strategy proposed in Florescu and Perkins, 2016; that is, performing the spectral method on the matrix $\mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top}\right)$. In this case, we have

$$\mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top}\right) - \boldsymbol{T}_i\boldsymbol{T}_i^{\top} = \underbrace{\mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top} - \boldsymbol{T}_i\boldsymbol{T}_i^{\top}\right)}_{\text{unbiased}} - \underbrace{\mathcal{P}_{\text{diag}}\left(\boldsymbol{T}_i\boldsymbol{T}_i^{\top}\right)}_{\text{diagonal deletion}}.$$

Therefore, if the diagonal elements of $\boldsymbol{T}_i\boldsymbol{T}_i^{\top}$ are not too large, the matrix $\mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top}\right)$ serves as a nearly unbiased estimator of $\boldsymbol{T}_i\boldsymbol{T}_i^{\top}$. It implies that the top-r eigenvectors of $\mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i\widehat{\boldsymbol{T}}_i^{\top}\right)$ could form a reliable estimator of the principal subspace of $\boldsymbol{T}_i\boldsymbol{T}_i^{\top}$. The complete description of the initialization procedure is summarized in Algorithm 3.

Remark 4 *The diagonal deletion idea has already been used in various scenarios, including low rank tensor completion (Cai et al., 2021b; Tong et al., 2021). We give a brief introduction here for the paper to be self-contained. In addition to diagonal deletion, one might also* *consider properly reweighting the diagonal entries, see for example (Cai and Zhang, 2016; Cho et al., 2017; Elsener and van de Geer, 2019; Loh and Wainwright, 2012; Lounici, 2013, 2014; Montanari and Sun, 2018; Zhang et al., 2018; Zhu et al., 2019).*

2.3 Implicit Regularization and Entrywise Convergence of RGM

Since the target tensor $\mathcal{T}$ is low rank, the application of HOSVD with the true parameter $\boldsymbol{r} = (r, r, r)$ yields the exact Tucker decomposition of $\mathcal{T}$, denoted

$$\mathcal{T} = \mathcal{S} \mathop{\times}_{i=1}^{3} \boldsymbol{U}_i,$$

where $\boldsymbol{U}_i \in \mathbb{R}^{n\times r}$ are the top-r left singular vectors of $\mathcal{M}_i(\mathcal{T})$ for $i = 1, 2, 3$ and $\mathcal{S} = \mathcal{T} \mathop{\times}_{i=1}^{3} \boldsymbol{U}_i^\mathsf{T} \in \mathbb{R}^{r\times r\times r}$ is the core tensor. As in matrix completion, the notion of incoherence is required for us to be able to successfully fill in the missing entries of a low rank tensor. Specifically, the incoherence parameter of $\mathcal{T}$ is defined as

$$\mu := \frac{n}{r} \max\left\{ \|\boldsymbol{U}_1\|_{2,\infty}^2, \|\boldsymbol{U}_2\|_{2,\infty}^2, \|\boldsymbol{U}_3\|_{2,\infty}^2 \right\}.$$

Clearly, the smallest value for μ can be 1, while the largest possible value is n/r. A tensor is μ-incoherent with a small μ implies that the singular vectors of its matricization form are weakly correlated with the canonical basis. Therefore, the energy of the tensor sufficiently spreads out across the measurement basis, and a small random subset of its entries still contains enough information for successful reconstruction. The following lemma follows directly from the definition of incoherence, see Section B.1 for the proof.

Lemma 5 *Letting $\boldsymbol{T}_i = \mathcal{M}_i(\mathcal{T})$, we have*

$$\|\boldsymbol{T}_i\|_{2,\infty} \le \sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \quad \left\|\boldsymbol{T}_i^\mathsf{T}\right\|_{2,\infty} \le \frac{\mu r}{n}\sigma_{\max}(\mathcal{T}), \quad \|\mathcal{T}\|_\infty \le \left(\frac{\mu r}{n}\right)^{3/2}\sigma_{\max}(\mathcal{T}).$$

The convergence of the Riemannian gradient method can be approximately decomposed into two phases. In Phase I, the iterates are not sufficiently close to the target tensor, we need to explicitly show that they remain in the incoherence region based on induction. In Phase II, the iterates are in a local neighbourhood of the target tensor where the restricted isometry property uniformly holds, and thus implicit regularization and entrywise convergence follows directly from the convergence in terms of the Frobenius norm.

Theorem 6 (Phase I convergence) *Suppose that $\mathcal{T}$ is μ-incoherent and the index set Ω satisfies the Bernoulli model with parameter p. If $n \ge C_0\kappa^6\mu^3 r^5$ and*

$$p \ge \max\left\{ \frac{C_1\kappa^8\mu^{3.5}r^6\log^3 n}{n^{3/2}}, \frac{C_2\kappa^{16}\mu^7 r^{12}\log^5 n}{n^2} \right\}$$

for some universal constants C_0, C_1 and C_2. Suppose $t_0 = 2\log_2 n + c$. Then with high probability, the iterates of RGM (Algorithm 2) satisfy

$$\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{U}_i\right\|_{2,\infty} \le \left(\frac{1}{2}\right)^t \sqrt{\frac{\mu r}{n}}, \quad i = 1, 2, 3,$$

$$\left\|\mathcal{X}^t - \mathcal{T}\right\|_\infty \leq \left(\frac{1}{2}\right)^t \frac{1}{n^{3/2}} \sigma_{\max}\left(\mathcal{T}\right), \tag{9}$$

for $t = 1, \cdots, t_0$, where $\boldsymbol{R}_i^t = \arg\min_{\boldsymbol{R}^\top \boldsymbol{R} = \boldsymbol{I}} \left\|\boldsymbol{X}_i^t \boldsymbol{R} - \boldsymbol{U}_i\right\|_{\mathsf{F}}$.

Theorem 6 establishes the performance guarantee of RGM within the first logarithm number of iterations, where $c > 0$ in t_0 is a constant specified in Remark 8. Analyzing the distance between factor matrices $\boldsymbol{X}_i^t$ and $\boldsymbol{U}_i$ is the key to showing the convergence of the RGM in phase I. As already mentioned, this explicitly shows that RGM automatically forces the iterates to stay incoherent in each iteration.

Theorem 7 (Phase II convergence) *Suppose that $\mathcal{T}$ is μ-incoherent and the index set Ω satisfies the Bernoulli model with parameter p. Suppose $p \geq \frac{C_2 \mu^2 r^2 \log n}{\varepsilon \cdot n^2}$ and*

$$\frac{\left\|\mathcal{X}^{t_0} - \mathcal{T}\right\|_{\mathsf{F}}}{\sigma_{\min}\left(\mathcal{T}\right)} \leq \frac{p^{1/2}\varepsilon}{9\left(1+\epsilon\right)},$$

where $0 < \varepsilon \leq \frac{1}{200}$. Then with high probability, the iterates of RGM (Algorithm 2) satisfy

$$\left\|\mathcal{X}^t - \mathcal{T}\right\|_{\mathsf{F}} \leq \left(\frac{1}{2}\right)^{t-t_0} \left\|\mathcal{X}^{t_0} - \mathcal{T}\right\|_{\mathsf{F}}, \quad \text{for all } t \geq t_0. \tag{10}$$

Remark 8 *Note that if we set $c = \lceil \log_2(9(1+\varepsilon)/\varepsilon) \rceil$ in Theorem 6, after Phase I, the iterates of RGM will enter the local neighborhood specified in Theorem 7. This follows from a simple calculation,*

$$\begin{aligned} \frac{\left\|\mathcal{X}^{t_0} - \mathcal{T}\right\|_{\mathsf{F}}}{\sigma_{\min}(\mathcal{T})} &\leq \frac{1}{\sigma_{\min}(\mathcal{T})} n^{3/2} \left\|\mathcal{X}^{t_0} - \mathcal{T}\right\|_\infty \overset{(a)}{\leq} \frac{1}{\sigma_{\min}(\mathcal{T})} n^{3/2} \cdot \frac{1}{2^{t_0}} \frac{1}{n^{3/2}} \sigma_{\max}(\mathcal{T}) \\ &\leq \frac{\kappa}{2^{t_0}} \overset{(b)}{\leq} \frac{\varepsilon}{9(1+\varepsilon)} \frac{\kappa}{n} \overset{(c)}{\leq} \frac{\sqrt{p}\varepsilon}{9(1+\varepsilon)}, \end{aligned}$$

where (a) is due to (9), (b) follows from $t_0 = 2\log_2 n + \lceil \log_2(9(1+\varepsilon)/\varepsilon) \rceil$, and (c) uses the fact $p \geq \kappa^2/n^2$.

As an immediate consequence of Theorem 6 and Theorem 7, one can obtain the proof of Theorem 1.

Proof [Proof of Theorem 1] Noting that the results in Theorem 6 naturally show that the inequalities (2) and (3) hold for $1 \leq t \leq t_0$. It only remains to show the above two inequalities hold for $t \geq t_0$.

A simple calculation yields that

$$\begin{aligned} \left\|\mathcal{X}^t - \mathcal{T}\right\|_\infty \leq \left\|\mathcal{X}^t - \mathcal{T}\right\|_{\mathsf{F}} &\leq \left(\frac{1}{2}\right)^{t-t_0} \left\|\mathcal{X}^{t_0} - \mathcal{T}\right\|_{\mathsf{F}} \leq \left(\frac{1}{2}\right)^{t-t_0} n^{3/2} \left\|\mathcal{X}^{t_0} - \mathcal{T}\right\|_\infty \\ &\leq \left(\frac{1}{2}\right)^{t-t_0} n^{3/2} \left(\frac{1}{2}\right)^{t_0} \frac{1}{n^{3/2}} \sigma_{\max}\left(\mathcal{T}\right) = \left(\frac{1}{2}\right)^t \sigma_{\max}\left(\mathcal{T}\right). \end{aligned}$$

Only detailed proofs for the $i = 1$ case are provided for (2), and the proofs for the other two cases are overall similar. Applying the Weyl's inequality yields that

$$\begin{aligned}\sigma_{\min}\left(\mathcal{M}_1\left(\mathcal{X}^t\right)\right) &\geq \sigma_{\min}\left(\mathcal{M}_1\left(\mathcal{T}\right)\right) - \left\|\mathcal{M}_1\left(\mathcal{X}^t\right) - \mathcal{M}_1\left(\mathcal{T}\right)\right\| \\ &\geq \sigma_{\min}\left(\mathcal{T}\right) - \left\|\mathcal{M}_1\left(\mathcal{X}^t\right) - \mathcal{M}_1\left(\mathcal{T}\right)\right\|_{\mathsf{F}} \\ &\geq \sigma_{\min}\left(\mathcal{T}\right) - \left(\frac{1}{2}\right)^{t-t_0}\left\|\mathcal{X}^{t_0} - \mathcal{T}\right\|_{\mathsf{F}} \\ &\geq \sigma_{\min}\left(\mathcal{T}\right) - \left(\frac{1}{2}\right)^{t-t_0}\frac{\sqrt{p}\varepsilon}{9(1+\varepsilon)}\sigma_{\min}\left(\mathcal{T}\right) \geq \frac{15}{16}\sigma_{\min}\left(\mathcal{T}\right),\end{aligned}$$

where the last inequality is due to $p \leq 1$ and $\varepsilon \leq \frac{1}{200}$. For any ℓ satisfies $1 \leq \ell \leq n$, using the fact $\mathcal{M}_1\left(\mathcal{X}^t\right) = \boldsymbol{X}_1^t \mathcal{M}_1\left(\mathcal{G}^t\right)\left(\boldsymbol{X}_3^t \otimes \boldsymbol{X}_2^t\right)^{\mathsf{T}}$, one can obtain

$$\begin{aligned}\left\|\boldsymbol{e}_\ell^{\mathsf{T}}\boldsymbol{X}_1^t\right\| &= \left\|\boldsymbol{e}_\ell^{\mathsf{T}}\mathcal{M}_1\left(\mathcal{X}^t\right)\left(\boldsymbol{X}_3^t \otimes \boldsymbol{X}_2^t\right)\mathcal{M}_1^{\dagger}\left(\mathcal{G}^t\right)\right\| \\ &\overset{(a)}{\leq} \left\|\boldsymbol{e}_\ell^{\mathsf{T}}\mathcal{M}_1\left(\mathcal{X}^t\right)\right\|\left\|\mathcal{M}_1^{\mathsf{T}}\left(\mathcal{G}^t\right)\left(\mathcal{M}_1\left(\mathcal{G}^t\right)\mathcal{M}_1^{\mathsf{T}}\left(\mathcal{G}^t\right)\right)^{-1}\right\| \\ &\leq \left(\left\|\boldsymbol{e}_\ell^{\mathsf{T}}\mathcal{M}_1\left(\mathcal{X}^t - \mathcal{T}\right)\right\| + \left\|\boldsymbol{e}_\ell^{\mathsf{T}}\mathcal{M}_1\left(\mathcal{T}\right)\right\|\right)\frac{1}{\sigma_{\min}\left(\mathcal{M}_1\left(\mathcal{G}^t\right)\right)} \\ &\leq \left(n\left\|\mathcal{X}^t - \mathcal{T}\right\|_\infty + \left\|\mathcal{M}_1\left(\mathcal{T}\right)\right\|_{2,\infty}\right)\frac{1}{\sigma_{\min}\left(\mathcal{M}_1\left(\mathcal{X}^t\right)\right)} \\ &\leq \left(n\left(\frac{1}{2}\right)^{t-t_0}\left(\frac{1}{2}\right)^{t_0}\sigma_{\max}\left(\mathcal{T}\right) + \sqrt{\frac{\mu r}{n}}\sigma_{\max}\left(\mathcal{T}\right)\right)\frac{16}{15\sigma_{\min}\left(\mathcal{T}\right)} \\ &\overset{(b)}{\leq} \left(n\left(\frac{1}{2}\right)^{t-t_0}\left(\frac{1}{2}\right)^{c}\frac{1}{n^2}\sigma_{\max}\left(\mathcal{T}\right) + \sqrt{\frac{\mu r}{n}}\sigma_{\max}\left(\mathcal{T}\right)\right)\frac{16}{15\sigma_{\min}\left(\mathcal{T}\right)} \leq 2\kappa\sqrt{\frac{\mu r}{n}},\end{aligned}$$

where (a) is due to $\sigma_{\min}\left(\mathcal{M}_1\left(\mathcal{G}^t\right)\right) > 0$ and (b) follows from $t_0 = 2\log_2 n + c$. ■

2.4 Numerical Experiments

We first test the convergence of RGM for tensor completion problem under two metrics defined below:

$$\textbf{Relative } \|\cdot\|_\infty\textbf{:} \quad \frac{\left\|\mathcal{X}^t - \mathcal{T}\right\|_\infty}{\|\mathcal{T}\|_\infty} \quad \text{and} \quad \textbf{Relative } \|\cdot\|_{2,\infty}\text{:} \quad \frac{\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{U}_i\right\|_{2,\infty}}{\|\boldsymbol{U}_i\|_{2,\infty}},$$

where $\boldsymbol{R}_i^t = \arg\min_{\boldsymbol{R}^{\mathsf{T}}\boldsymbol{R}=\boldsymbol{I}}\left\|\boldsymbol{X}_i^t\boldsymbol{R} - \boldsymbol{U}_i\right\|_{\mathsf{F}}$. Tests are conducted with $n = 100$, $r = 3$ and $p = 0.2$. For fixed (n, r), test tensors are generated through the Tucker decomposition $\mathcal{T} = \mathcal{S} \overset{3}{\underset{i=1}{\times}} \boldsymbol{U}_i$, where $\mathcal{S} \in \mathbb{R}^{r\times r\times r}$ is a random tensor with i.i.d. entries of standard normal distribution and $\{\boldsymbol{U}_i\}_{i=1}^3$ are random orthonormal matrices of size $n \times r$ which are obtained via the orthogonalization of standard Gaussian matrices. The plot of average relative errors over 100 random tests against iteration count is presented in Figure 1 (Right). Clearly, a

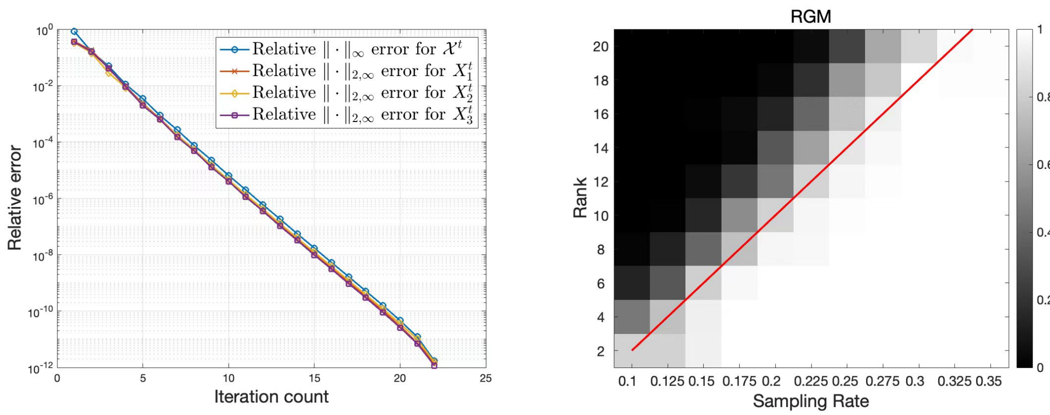

Figure 1: (Left) Relative errors of $\mathcal{X}^t$ and $\boldsymbol{X}_i^t$ for $i = 1, 2, 3$ against iteration count when $n = 100$, $r = 3$, $p = 0.2$; (Right) Phase transition plot for varying r and sampling rate when $n = 100$ (different colors represent different successful rates out of 100 random tests).

desirable linear convergence can be observed from the plots for both the iterates and the factor matrices.

Our results suggests that $O(n^{3/2}r^6)$ are overall sufficient for the successful reconstruction of a low Tucker-rank tensor. For the dependence on n, it matches the lower bound for any known polynomial time algorithms. Moreover, it is shown in Barak and Moitra, 2016 that, conditioned on some conjecture on computational complexity theory, no polynomial time algorithm can be successful if the sampling complexity is less than $O(n^{3/2})$. For the dependence on r, we conduct the phase transition tests for fixed $n = 100$, and varying r and p, see Figure 1 (Right). For each pair of (r, p), 100 random random trials are tested and a trial is considered to have successfully recover the target tensor if the output tensor $\mathcal{X}^t$ satisfies $\frac{\|\mathcal{X}^t - \mathcal{T}\|_\infty}{\|\mathcal{T}\|_\infty} \leq 10^{-3}$. The phase transition plot indicates that the sampling complexity for successful recovery is about linearly proportional to r, which suggests the possibility of reducing this dependency in the future.

3. Proof of Theorem 6

The proof of Theorem 6 relies on the reconstruction of auxiliary sequences via a *leave-one-out* perturbation argument, and is much more involved. Thus, this section is devoted to the proof outline of Theorem 6, while the proofs of the intermediate results are deferred to later sections.

To obtain the entrywise error bound of the iterates, the $\ell_{2,\infty}$ norm of the factor matrices needs to be bounded which is quite difficult to control directly due to the complicated statistical dependency. To overcome this difficulty, the leave-one-out technique proposes to introduce a collection of leave-one-out versions of $\{\mathcal{X}^t\}$, denoted by $\{\mathcal{X}^{t,\ell}\}$ for each

$1 \leq \ell \leq n$. Specifically, for every ℓ, define the following auxiliary loss function

$$\frac{1}{2p}\left\|\mathcal{P}_{\Omega_{-\ell}}\left(\mathcal{X}-\mathcal{T}\right)\right\|_{\mathsf{F}}^2+\frac{1}{2}\left\|\mathcal{P}_{\ell}\left(\mathcal{X}-\mathcal{T}\right)\right\|_{\mathsf{F}}^2, \tag{11}$$

where $\mathcal{P}_{\Omega_{-\ell}}\left(\mathcal{Z}\right)$ and $\mathcal{P}_{\ell}\left(\mathcal{Z}\right)$ are defined as follows:

$$\left[\mathcal{P}_{\Omega_{-\ell}}(\mathcal{Z})\right]_{i_1,i_2,i_3} := \begin{cases}[\mathcal{Z}]_{i_1,i_2,i_3}, & \text{if } i_1\neq\ell, i_2\neq\ell, i_3\neq\ell \text{ and } (i_1,i_2,i_3)\in\Omega,\\ 0, & \text{otherwise,}\end{cases}$$
$$\left[\mathcal{P}_{\ell}(\mathcal{Z})\right]_{i_1,i_2,i_3} := \begin{cases}[\mathcal{Z}]_{i_1,i_2,i_3}, & \text{if } i_1=\ell, \text{or } i_2=\ell, \text{or } i_3=\ell,\\ 0, & \text{otherwise.}\end{cases}$$

The leave-one-out sequence $\left\{\mathcal{X}^{t,\ell}\right\}_{t\geq 1}$ is produced by applying RGM to this new cost function. If Ω satisfies the Bernoulli model, then we can rewrite (11) as

$$\frac{1}{2p}\sum_{i_1,i_2,i_3\neq\ell}\delta_{i_1,i_2,i_3}\left(\mathcal{X}_{i_1,i_2,i_3}-\mathcal{T}_{i_1,i_2,i_3}\right)^2+\frac{1}{2}\sum_{\exists i_j=\ell,\ j\in[3]}\left(\mathcal{X}_{i_1,i_2,i_3}-\mathcal{T}_{i_1,i_2,i_3}\right)^2, \tag{12}$$

where $\{\delta_{i_1,i_2,i_3}\}$ are n^3 independent Bernoulli random variables. Noting that (12) does not depend on $\{\delta_{i_1,i_2,i_3}:\ \exists i_j=\ell,\ j\in[3]\}$, the sequence $\left\{\mathcal{X}^{t,\ell}\right\}_{t\geq 1}$ is independent of those random variables provided the initial guess $\mathcal{X}^{1,\ell}$ is independent of them. This decoupling of the statistical dependency turns out to be crucial for us to bound the $\ell_{2,\infty}$ norm of the factor matrices. The initial guess $\mathcal{X}^{1,\ell}$ can be similarly generated by the spectral method with diagonal deletion, but with those entries at locations indexed by $\{(i_1,i_2,i_3):\ \exists i_j=\ell,\ j\in[3]\}$ being replaced by the ground truth values. The complete procedure to create the leave-one-out sequence $\left\{\mathcal{X}^{t,\ell}\right\}_{t\geq 1}$ is described in Algorithm 4. We would like to caution that Algorithm 4 is by no means a practical algorithm, but only introduced for the sake of analysis.

To facilitate the analysis, it is much more convenient to rewrite the iterates of Algorithm 2 into the following perturbation form,

$$\mathcal{X}^{t+1}=\mathcal{H}_{\boldsymbol{r}}\left(\mathcal{T}+\mathcal{E}^t\right),$$

where the residual tensor $\mathcal{E}^t$ is given by

$$\mathcal{E}^t := \left(\mathcal{I}-p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_{\Omega}\right)\left(\mathcal{X}^t-\mathcal{T}\right),\quad t\geq 1. \tag{13}$$

For $t=0$, since $\mathcal{X}^1$ can be rewritten as

$$\mathcal{X}^1=\mathcal{H}_{\boldsymbol{r}}\left(\mathcal{T}+\mathcal{X}^1-\mathcal{T}\right),$$

the residual tensor in the initialization step is defined as

$$\mathcal{E}^0 := \mathcal{X}^1-\mathcal{T}=\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega}\right)(-\mathcal{T})\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^1\boldsymbol{X}_i^{1\mathsf{T}}+\mathcal{T}\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^1\boldsymbol{X}_i^{1\mathsf{T}}-\mathcal{T}. \tag{14}$$

Algorithm 4 The ℓ-th leave-one-out sequence for tensor completion

Input: tensors $\mathcal{P}_{\Omega_{-\ell}}(\mathcal{T})$, $\mathcal{P}_{\ell}(\mathcal{T})$, multilinear rank $\boldsymbol{r}=(r,r,r)$, parameter p.
for $i=1,2,3$ **do**
    Let $\boldsymbol{X}_i^{1,\ell}\boldsymbol{\Sigma}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell^\top}$ be the top-r eigenvalue decomposition of $\mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i^{\ell}\widehat{\boldsymbol{T}}_i^{\ell^\top}\right)$, where $\widehat{\boldsymbol{T}}_i^{\ell}=\mathcal{M}_i\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}}(\mathcal{T})+\mathcal{P}_{\ell}(\mathcal{T})\right)$.
end for
$\mathcal{G}^{1,\ell}=\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}}(\mathcal{T})+\mathcal{P}_{\ell}(\mathcal{T})\right)\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i^{1,\ell^\top}$.
$\mathcal{X}^{1,\ell}=\mathcal{G}^{1,\ell}\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i^{1,\ell}$.
for $t=1,\cdots$ **do**
    for $i=1,2,3$ **do**
        $\boldsymbol{X}_i^{t+1,\ell}=\mathrm{SVD}_r\left(\mathcal{M}_i\left(\mathcal{X}^{t,\ell}-\mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}}+\mathcal{P}_{\ell}\right)\left(\mathcal{X}^{t,\ell}-\mathcal{T}\right)\right)\right)$.
    end for
    $\mathcal{G}^{t+1,\ell}=\left(\mathcal{X}^{t,\ell}-\mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}}+\mathcal{P}_{\ell}\right)\left(\mathcal{X}^{t,\ell}-\mathcal{T}\right)\right)\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i^{t+1,\ell^\top}$.
    $\mathcal{X}^{t+1,\ell}=\mathcal{G}^{t+1,\ell}\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i^{t+1,\ell}$.
end for

Similarly, the residual tensors of the ℓ-th leave one out sequence from Algorithm 4 are defined by

$$\mathcal{E}^{0,\ell}:=\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega_{-\ell}}-\mathcal{P}_{\ell}\right)(-\mathcal{T})\right)\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell^\top}+\mathcal{T}\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell^\top}-\mathcal{T}, \tag{15}$$

$$\mathcal{E}^{t,\ell}:=\left(\mathcal{I}-\mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}}+\mathcal{P}_{\ell}\right)\right)\left(\mathcal{X}^{t,\ell}-\mathcal{T}\right),\quad t\geq 1, \tag{16}$$

which satisfy

$$\mathcal{X}^{1,\ell}=\mathcal{H}_{\boldsymbol{r}}\left(\mathcal{T}+\mathcal{E}^{0,\ell}\right)\quad\text{and}\quad\mathcal{X}^{t+1,\ell}=\mathcal{H}_{\boldsymbol{r}}\left(\mathcal{T}+\mathcal{E}^{t,\ell}\right).$$

Let $\boldsymbol{T}_i=\mathcal{M}_i(\mathcal{T})$, $\boldsymbol{E}_i^{t-1}=\mathcal{M}_i\left(\mathcal{E}^{t-1}\right)$ and $\boldsymbol{E}_i^{t-1,\ell}=\mathcal{M}_i\left(\mathcal{E}^{t-1,\ell}\right)$ be the mode-i matricizations of the corresponding tensors. It can be seen from Algorithms 2 and 4 that the matrices $\boldsymbol{X}_i^{t}$ and $\boldsymbol{X}_i^{t,\ell}$ are the top-r eigenvectors of $\left(\boldsymbol{T}_i+\boldsymbol{E}_i^{t-1}\right)\left(\boldsymbol{T}_i+\boldsymbol{E}_i^{t-1}\right)^\top$ and $\left(\boldsymbol{T}_i+\boldsymbol{E}_i^{t-1,\ell}\right)\left(\boldsymbol{T}_i+\boldsymbol{E}_i^{t-1,\ell}\right)^\top$, respectively. Recall that the eigenvalue decomposition of $\boldsymbol{T}_i\boldsymbol{T}_i^\top$ is $\boldsymbol{T}_i\boldsymbol{T}_i^\top=\boldsymbol{U}_i\boldsymbol{\Lambda}_i\boldsymbol{U}_i^\top$. If we further define three auxiliary $r\times r$ orthonormal matrices as follows:

$$\boldsymbol{R}_i^{t}=\arg\min_{\boldsymbol{R}}\left\|\boldsymbol{X}_i^{t}\boldsymbol{R}-\boldsymbol{U}_i\right\|_{\mathsf{F}},\quad \boldsymbol{R}_i^{t,\ell}=\arg\min_{\boldsymbol{R}}\left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{R}-\boldsymbol{U}_i\right\|_{\mathsf{F}},$$
$$\boldsymbol{T}_i^{t,\ell}=\arg\min_{\boldsymbol{R}}\left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{R}-\boldsymbol{X}_i^{t}\boldsymbol{R}_i^{t}\right\|_{\mathsf{F}},$$

the following theorem for the sequences produced by Algorithms 2 and 4.

Theorem 9 *Under the assumption of Theorem 6, the following inequalities hold with high probability for all* $1\leq\ell\leq n$ *and* $1\leq t\leq t_0$,

$$\left\|\boldsymbol{E}_i^{t-1}\right\|\leq\frac{1}{2^{20}\kappa^6\mu^2r^4}\frac{1}{2^t}\sigma_{\max}(\mathcal{T}), \tag{17a}$$

$$\left\| \boldsymbol{E}_i^{t-1,\ell} \right\| \le \frac{1}{2^{20}\kappa^6\mu^2 r^4}\frac{1}{2^t}\sigma_{\max}(\mathcal{T}), \tag{17b}$$

$$\left\| \boldsymbol{X}_i^{t,\ell}\boldsymbol{R}_i^{t,\ell} - \boldsymbol{U}_i \right\|_{2,\infty} \le \frac{1}{2^{20}\kappa^2\mu^2 r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}, \tag{17c}$$

$$\left\| \mathcal{E}^{t-1} - \mathcal{E}^{t-1,\ell} \right\|_{\mathsf{F}} \le \frac{1}{2^{20}\kappa^4\mu^2 r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \tag{17d}$$

$$\left\| \boldsymbol{X}_i^{t}\boldsymbol{R}_i^{t} - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell} \right\|_{\mathsf{F}} \le \frac{1}{2^{20}\kappa^2\mu^2 r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}, \tag{17e}$$

$$\left\| \boldsymbol{X}_i^{t}\boldsymbol{R}_i^{t} - \boldsymbol{U}_i \right\|_{2,\infty} \le \frac{1}{2^{20}\kappa^2\mu^2 r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}. \tag{17f}$$

We also need a lemma which transfers the convergence result in terms of the ℓ_2 and $\ell_{2,\infty}$ norms to that in terms of the ℓ_∞ norm.

Lemma 10 *Let $\mathcal{X} = \mathrm{HOSVD}_r(\mathcal{T}+\mathcal{E})$ for some perturbation tensor $\mathcal{E} \in \mathbb{R}^{n\times n\times n}$. Let the Tucker decomposition of $\mathcal{X}$ be $\mathcal{G} \times_{i=1}^{3} \boldsymbol{X}_i$ with $\boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{X}_i = \boldsymbol{I}$. Define*

$$\boldsymbol{R}_i = \arg\min_{\boldsymbol{R}^{\mathsf{T}}\boldsymbol{R}=\boldsymbol{I}} \left\|\boldsymbol{X}_i\boldsymbol{R} - \boldsymbol{U}_i\right\|_{\mathsf{F}} \text{ and } B = \max_{i=1,2,3}\left(\left\|\boldsymbol{X}_i\boldsymbol{R}_i - \boldsymbol{U}_i\right\|_{2,\infty} + \frac{15\sigma_{\max}(\mathcal{T})}{2\sigma_{\min}^2(\mathcal{T})}\left\|\boldsymbol{U}_i\right\|_{2,\infty}\left\|\boldsymbol{E}_i\right\| \right),$$

where $\boldsymbol{E}_i := \mathcal{M}_i(\mathcal{E})$. Suppose that $\max_{i=1,2,3}\|\boldsymbol{E}_i\| \le \sigma_{\max}(\mathcal{T})/\left(10\kappa^2\right)$. Then one has

$$\|\mathcal{X}-\mathcal{T}\|_\infty \le \sigma_{\max}(\mathcal{T})\left(B^3 + 3B^2\left(\max_{i=1,2,3}\|\boldsymbol{U}_i\|_{2,\infty}\right) + 3B\left(\max_{i=1,2,3}\|\boldsymbol{U}_i\|_{2,\infty}\right)^2 \right) + \max_{i=1,2,3}\|\boldsymbol{E}_i\|\prod_{i=1}^{3}\|\boldsymbol{X}_i\|_{2,\infty}.$$

Theorem 9 will be proved by induction, with the proof details for the base case and induction step presented in Sections A.1 and A.2, respectively. The proof of Lemma 10 can be found in Section B.2. Equipped with Theorem 9 and Lemma 10, we are now able to present the proof of Theorem 6.

Proof [Proof of Theorem 6] The $\ell_{2,\infty}$ convergence for the factor matrices follows directly from (17f). It only remains to show the ℓ_∞ convergence of the iterates. From (17a), one can see that

$$\left\|\boldsymbol{E}_i^{t-1}\right\| \le \frac{1}{10\kappa^2}\sigma_{\max}(\mathcal{T}), \quad i=1,2,3.$$

Applying Lemma 10 with $\mathcal{X} := \mathcal{X}^t$, $\mathcal{E} := \mathcal{E}^{t-1}$ yields that

$$\left\|\mathcal{X}^t-\mathcal{T}\right\|_\infty \le \sigma_{\max}(\mathcal{T})\left(B^3 + 3B^2\left(\max_{i=1,2,3}\|\boldsymbol{U}_i\|_{2,\infty}\right) + 3B\left(\max_{i=1,2,3}\|\boldsymbol{U}_i\|_{2,\infty}\right)^2 \right) + \max_{i=1,2,3}\left\|\boldsymbol{E}_i^{t-1}\right\|\prod_{i=1}^{3}\left\|\boldsymbol{X}_i^t\right\|_{2,\infty}. \tag{18}$$

By (17a) and (17f), one has

$$B \leq \frac{1}{2^{16}\kappa^2\mu^2 r^4}\sqrt{\frac{\mu r}{n}}. \tag{19}$$

Finally, substituting (17a) and (19) into (18) yields that

$$\left\|\mathcal{X}^t - \mathcal{T}\right\|_\infty \leq \left(\frac{1}{2}\right)^t \frac{1}{n^{3/2}}\sigma_{\max}\left(\mathcal{T}\right). \tag{20}$$

Noting that (20) holds with high probability for each t and $t_0 = 2\log_2 n + c$. ■

4. Proof of Theorem 7

In this section, we prove Theorem 7 by induction. It is trivial that when $t = t_0$ the result holds. Next, we assume (10) holds for the iterations $t_0, t_0+1, \cdots, t$, and then prove it also holds for $t+1$. The proof relies on several lemmas which reveal the uniform restricted isometry property in a local neighborhood of the ground truth. We first present these lemmas and postpone the proofs to Section B.

Lemma 11 *Suppose T is the tangent space of $\mathbb{M}_r$ at $\mathcal{T}$. Then*

$$\|\mathcal{P}_T(\boldsymbol{e}_{i_1} \circ \boldsymbol{e}_{i_2} \circ \boldsymbol{e}_{i_3})\|_{\mathsf{F}}^2 \leq 4\left(\frac{\mu r}{n}\right)^2. \tag{21}$$

Lemma 12 *Suppose Ω is sampled according to the Bernoulli model and the tensor $\mathcal{T}$ obeys the incoherence condition with parameter μ. If $p \geq \frac{C_2\mu^2 r^2 \log n}{\varepsilon \cdot n^2}$, then*

$$\left\|\mathcal{P}_T\left(p^{-1}\mathcal{P}_\Omega - \mathcal{I}\right)\mathcal{P}_T\right\| \leq \varepsilon$$

holds with high probability, where $\varepsilon > 0$ is an absolute constant.

Remark 13 *Lemma 11 and Lemma 12 highly resemble Lemma 2 in Yuan and Zhang, 2016 and Lemma 12 in Tong et al., 2021, respectively. However, the definition of the operator $\mathcal{P}_T$ in this paper differs from the one in Tong et al., 2021; Yuan and Zhang, 2016. Specifically, the operator $\mathcal{P}_T$ in the previous works was defined as*

$$\mathcal{P}_T\left(\mathcal{Z}\right) = \mathcal{Z} \mathop{\times}_{i=1}^{3} \boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}} + \sum_{i=1}^{3} \mathcal{Z} \times_i \left(\boldsymbol{I} - \boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\right) \mathop{\times}_{j\neq i} \boldsymbol{U}_j\boldsymbol{U}_j^{\mathsf{T}}.$$

Lemma 14 *Let $\mathcal{X}^t = \mathcal{G}^t \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i^t$ and $\mathcal{T} = \mathcal{S} \mathop{\times}_{i=1}^{3} \boldsymbol{U}_i$ be two tensors in $\mathbb{M}_r$ and $T_{\mathcal{X}^t}, T$ be the tangent spaces of $\mathbb{M}_r$ at $\mathcal{X}^t$ and $\mathcal{T}$, respectively. Then*

$$\left\|\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}} - \boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\right\| \leq \frac{1}{\sigma_{\min}(\mathcal{T})}\left\|\mathcal{X}^t - \mathcal{T}\right\|_{\mathsf{F}}, \quad i = 1,2,3,$$
$$\left\|\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_T\right\| \leq \frac{9}{\sigma_{\min}(\mathcal{T})}\left\|\mathcal{X}^t - \mathcal{T}\right\|_{\mathsf{F}}.$$

Lemma 15 *Assume the following inequalities hold for* $0 < \varepsilon \leq \frac{1}{200}$,

$$\left\|\mathcal{P}_T - p^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T\right\| \leq \varepsilon \quad \text{and} \quad \frac{\left\|\mathcal{X}^t - \mathcal{T}\right\|_\mathsf{F}}{\sigma_{\min}(\mathcal{T})} \leq \frac{\sqrt{p}\varepsilon}{9(1+\varepsilon)}. \tag{22}$$

Then

$$\left\|\mathcal{P}_\Omega\mathcal{P}_{T_{\mathcal{X}^t}}\right\| \leq 2\sqrt{p}(1+\varepsilon) \quad \text{and} \quad \left\|\mathcal{P}_{T_{\mathcal{X}^t}} - p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_\Omega\mathcal{P}_{T_{\mathcal{X}^t}}\right\| \leq 5\varepsilon.$$

Proof [Proof of Theorem 7] First, Lemma 12 implies that

$$\left\|\mathcal{P}_T - p^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T\right\| \leq \varepsilon$$

holds with high probability provided that $p \geq \frac{C_2\mu^2 r^2\log n}{\varepsilon\cdot n^2}$.

We assume that in the t-th iteration $\mathcal{X}^t$ satisfies

$$\frac{\left\|\mathcal{X}^t - \mathcal{T}\right\|_\mathsf{F}}{\sigma_{\min}(\mathcal{T})} \leq \frac{\sqrt{p}\varepsilon}{9(1+\varepsilon)}. \tag{23}$$

Recall that $\mathcal{X}^{t+1} = \mathcal{H}_r\left(\mathcal{X}^t - p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_\Omega\left(\mathcal{X}^t - \mathcal{T}\right)\right) = \mathcal{H}_r\left(\mathcal{T} + \mathcal{E}^t\right)$. One can see that

$$\left\|\mathcal{X}^{t+1} - \mathcal{T}\right\|_\mathsf{F} \leq \left\|\mathcal{X}^{t+1} - \left(\mathcal{T} + \mathcal{E}^t\right)\right\|_\mathsf{F} + \left\|\mathcal{E}^t\right\|_\mathsf{F} \leq (\sqrt{3}+1)\left\|\mathcal{E}^t\right\|_\mathsf{F} \leq 2^2\left\|\mathcal{E}^t\right\|_\mathsf{F},$$

where the second inequality follows from the quasi-optimality of HOSVD. It remains to bound $\left\|\mathcal{E}^t\right\|_\mathsf{F}$. To this end, invoking the triangle inequality gives that

$$\begin{aligned}\left\|\mathcal{E}^t\right\|_\mathsf{F} &= \left\|\mathcal{X}^t - \mathcal{T} - p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_\Omega\left(\mathcal{X}^t - \mathcal{T}\right)\right\|_\mathsf{F} \\ &\leq \underbrace{\left\|\left(\mathcal{P}_{T_{\mathcal{X}^t}} - p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_\Omega\mathcal{P}_{T_{\mathcal{X}^t}}\right)\left(\mathcal{X}^t - \mathcal{T}\right)\right\|_\mathsf{F}}_{=:\upsilon_1} + \underbrace{\left\|\left(\mathcal{I} - \mathcal{P}_{T_{\mathcal{X}^t}}\right)\left(\mathcal{X}^t - \mathcal{T}\right)\right\|_\mathsf{F}}_{=:\upsilon_2} \\ &\quad + \underbrace{\left\|p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_\Omega\left(\mathcal{I} - \mathcal{P}_{T_{\mathcal{X}^t}}\right)\left(\mathcal{X}^t - \mathcal{T}\right)\right\|_\mathsf{F}}_{=:\upsilon_3}.\end{aligned}$$

Bounding υ_1. Applying Lemma 15 gives that

$$\upsilon_1 \leq \left\|\mathcal{P}_{T_{\mathcal{X}^t}} - p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_\Omega\mathcal{P}_{T_{\mathcal{X}^t}}\right\| \cdot \left\|\mathcal{X}^t - \mathcal{T}\right\|_\mathsf{F} \leq 5\varepsilon\left\|\mathcal{X}^t - \mathcal{T}\right\|_\mathsf{F}.$$

Bounding υ_2. The application of (Cai et al., 2020, Lemma 5.2) gives that

$$\upsilon_2 = \left\|\left(\mathcal{I} - \mathcal{P}_{T_{\mathcal{X}^t}}\right)(\mathcal{T})\right\|_\mathsf{F} \leq \frac{8}{\sigma_{\min}(\mathcal{T})}\left\|\mathcal{X}^t - \mathcal{T}\right\|_\mathsf{F}^2 \leq \frac{8\sqrt{p}\varepsilon}{9(1+\varepsilon)}\left\|\mathcal{X}^t - \mathcal{T}\right\|_\mathsf{F} \leq 2\varepsilon\left\|\mathcal{X}^t - \mathcal{T}\right\|_\mathsf{F}.$$

Bounding υ_3. A direct computation yields that

$$\begin{aligned}\upsilon_3 &\leq p^{-1}\left\|\mathcal{P}_\Omega\mathcal{P}_{T_{\mathcal{X}^t}}\right\| \cdot \left\|\left(\mathcal{I} - \mathcal{P}_{T_{\mathcal{X}^t}}\right)(\mathcal{T})\right\|_\mathsf{F} \\ &\leq p^{-1} \cdot 2\sqrt{p}(1+\varepsilon) \cdot \frac{8}{\sigma_{\min}(\mathcal{T})}\left\|\mathcal{X}^t - \mathcal{T}\right\|_\mathsf{F}^2 \leq 16\varepsilon\left\|\mathcal{X}^t - \mathcal{T}\right\|_\mathsf{F},\end{aligned}$$

where the second line is due to Lemma 15.

Putting the above bounds together yields that

$$\left\|\mathcal{X}^{t+1}-\mathcal{T}\right\|_{\mathsf{F}} \leq 4\left\|\mathcal{E}^{t}\right\|_{\mathsf{F}} \overset{(a)}{\leq} 100\varepsilon \cdot \left\|\mathcal{X}^{t}-\mathcal{T}\right\|_{\mathsf{F}} \leq \frac{1}{2}\left\|\mathcal{X}^{t}-\mathcal{T}\right\|_{\mathsf{F}}, \tag{24}$$

where (a) is due to the induction condition and the last step follows from $\varepsilon \leq \frac{1}{200}$.

By the assumption of the theorem, the inequality (23) is valid for $t = t_0$. Since $\left\|\mathcal{X}^{t}-\mathcal{T}\right\|_{\mathsf{F}}$ is a contractive sequence following from (24), the inequality (23) is valid for all $t \geq t_0$ by induction. ■

5. Conclusion and Discussion

In this paper, entrywise convergence of the vanilla Riemannian gradient method for low Tucker-rank tensor completion has been established and the implicit regularization property of the method has been revealed. For conciseness of presentation, we focus on three-way tensors and the noiseless case. Indeed, the results can be extended to general multi-way tensors and the noisy case, with the sampling complexity and error bound matching these results in Cai et al., 2021b; Xia and Yuan, 2019; Xia et al., 2021, see the supplement of this paper in Wang et al., 2021 for details. For future work, it is interesting to further optimize the dependency of the sampling complexity on the rank r. Additionally, it may also be possible to extend the analysis to the low rank tensor completion problem based on the tensor train decomposition since both the Tucker decomposition and the tensor train decomposition reduce to the same form of decomposition for the matrix case.

Acknowledgments

This work was partially supported by the National Key R&D Program of China (Grant No. 2021YFA1003300), Natural Science Foundation of Shanghai (Grant No. 23ZR1406400), and National Science Foundation of China (Grant No. 12001108).

Appendix A. Proof of Theorem 9

A.1 Base Case for Theorem 9

We first list two useful lemmas, whose proofs are deferred to Sections B.7 and B.8.

Lemma 16 *Under the assumption of Theorem 6, the following inequality holds with high probability,*

$$\begin{aligned}&\left\|\mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i^{\ell}\widehat{\boldsymbol{T}}_i^{\ell\mathsf{T}}\right)-\boldsymbol{T}_i\boldsymbol{T}_i^{\mathsf{T}}\right\|\\&\leq C\left(\left(\frac{\mu^{3/2}r^{3/2}}{n^{3/2}p}+\frac{\mu^2r^2}{n^2p}\right)\log n+\sqrt{r}\left(\frac{\mu^{3/2}r^{3/2}\log^3 n}{n^{3/2}p}+\sqrt{\frac{\mu^2r^2\log^5 n}{n^2p}}\right)+\frac{\mu r}{n}\right)\sigma_{\max}^2(\mathcal{T}).\end{aligned}$$

Lemma 17 *Suppose $n \geq C_0 \mu^3 r^5 \kappa^8$ and*

$$p \geq \max\left\{\frac{C_1 \kappa^8 \mu^{3.5} r^6 \log^3 n}{n^{3/2}}, \frac{C_2 \kappa^{16} \mu^6 r^{11} \log^5 n}{n^2}\right\}. \tag{25}$$

With high probability, one has

$$\left\|\boldsymbol{X}_i^1 \boldsymbol{R}_i^1 - \boldsymbol{U}_i\right\| \leq \frac{1}{2^5} \frac{1}{2^{20} \kappa^6 \mu^2 r^4}, \tag{26}$$

$$\left\|\boldsymbol{X}_i^{1,\ell} \boldsymbol{R}_i^{1,\ell} - \boldsymbol{U}_i\right\| \leq \frac{1}{2^5} \frac{1}{2^{20} \kappa^6 \mu^2 r^4}. \tag{27}$$

A.1.1 Bounding $\left\|\boldsymbol{E}_i^0\right\|$ and $\left\|\boldsymbol{E}_i^{0,\ell}\right\|$

In light of the symmetry of tensor matricization among different modes, we only focus on bounding $\left\|\boldsymbol{E}_1^0\right\|$ and $\left\|\boldsymbol{E}_1^{0,\ell}\right\|$, and omit the proofs for the other two modes. The spectral norm of $\boldsymbol{E}_1^0$ can be decomposed as follows:

$$\begin{aligned}\left\|\boldsymbol{E}_1^0\right\| &= \left\|\mathcal{M}_1\left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T}) \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top} + \mathcal{T} \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top} - \mathcal{T}\right)\right\| \\ &\leq \underbrace{\left\|\mathcal{M}_1\left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T}) \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top}\right)\right\|}_{=:\varphi_1} + \underbrace{\left\|\mathcal{M}_1\left(\mathcal{T} \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top} - \mathcal{T}\right)\right\|}_{=:\varphi_2}.\end{aligned}$$

Bounding φ_1. Notice that $\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T}) \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top}$ is a tensor of multilinear rank at most (r, r, r). Applying Lemma 22 yields that

$$\begin{aligned}\varphi_1 &= \left\|\mathcal{M}_1\left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T}) \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top}\right)\right\| \leq \sqrt{r}\left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T}) \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top}\right\| \\ &\leq \sqrt{r}\left\|\left(p^{-1}\mathcal{P}_\Omega - \mathcal{I}\right)(\mathcal{T})\right\| \overset{(a)}{\leq} \sqrt{r} C\left(\frac{\log^3 n}{p}\|\mathcal{T}\|_\infty + \sqrt{\frac{3\log^5 n}{p}} \max_{i=1,2,3}\left\|\boldsymbol{T}_i^\top\right\|_{2,\infty}^2\right) \\ &\overset{(b)}{\leq} \sqrt{r} C\left(\frac{\mu^{3/2} r^{3/2} \log^3 n}{n^{3/2} p} + \sqrt{\frac{\mu^2 r^2 \log^5 n}{n^2 p}}\right) \sigma_{\max}(\mathcal{T}) \overset{(c)}{\leq} \frac{1}{2} \frac{1}{2^{20} \kappa^6 \mu^2 r^4} \frac{1}{2} \sigma_{\max}(\mathcal{T}),\end{aligned} \tag{28}$$

where (a) is due to Lemma 23, (b) follows from Lemma 5, and (c) is due to the assumption $p \geq \max\left\{\frac{C_1 \kappa^6 \mu^{3.5} r^6 \log^3 n}{n^{3/2}}, \frac{C_2 \kappa^{12} \mu^6 r^{11} \log^5 n}{n^2}\right\}$.

Bounding φ_2. Since $\mathcal{T} = \mathcal{T} \mathop{\times}_{i=1}^{3} \boldsymbol{U}_i \boldsymbol{U}_i^\top$, one has

$$\begin{aligned}\mathcal{T} \mathop{\times}_{i=1}^{3} \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top} - \mathcal{T} &= \mathcal{T} \times_1 \left(\boldsymbol{X}_1^1 \boldsymbol{X}_1^{1\top} - \boldsymbol{U}_1 \boldsymbol{U}_1^\top\right) \times_2 \boldsymbol{X}_2^1 \boldsymbol{X}_2^{1\top} \times_3 \boldsymbol{X}_3^1 \boldsymbol{X}_3^{1\top} \\ &\quad + \mathcal{T} \times_1 \boldsymbol{U}_1 \boldsymbol{U}_1^\top \times_2 \left(\boldsymbol{X}_2^1 \boldsymbol{X}_2^{1\top} - \boldsymbol{U}_2 \boldsymbol{U}_2^\top\right) \times_3 \boldsymbol{X}_3^1 \boldsymbol{X}_3^{1\top}\end{aligned}$$

$$+ \mathcal{T} \times_1 \boldsymbol{U}_1 \boldsymbol{U}_1^\top \times_2 \boldsymbol{U}_2 \boldsymbol{U}_2^\top \times_3 \left(\boldsymbol{X}_3^1 \boldsymbol{X}_3^{1\top} - \boldsymbol{U}_3 \boldsymbol{U}_3^\top \right).$$

Consequently,

$$\begin{aligned}
\varphi_2 = \left\| \mathcal{M}_1 \left(\mathcal{T} \underset{i=1}{\overset{3}{\times}} \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top} - \mathcal{T} \right) \right\| &\leq 3 \max_{i=1,2,3} \left\| \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top} - \boldsymbol{U}_i \boldsymbol{U}_i^\top \right\| \cdot \max_{i=1,2,3} \|\boldsymbol{T}_i\| \\
&\leq 6 \max_{i=1,2,3} \left\| \boldsymbol{X}_i^1 \boldsymbol{R}_i^1 - \boldsymbol{U}_i \right\| \cdot \sigma_{\max}(\mathcal{T}) \overset{(a)}{\leq} 6 \cdot \frac{1}{2^{25} \kappa^6 \mu^2 r^4} \cdot \sigma_{\max}(\mathcal{T}), \qquad (29)
\end{aligned}$$

where (a) follows from (26).

Putting (28) and (29) together shows that

$$\left\| \boldsymbol{E}_1^0 \right\| \leq \frac{1}{2^{20} \kappa^6 \mu^2 r^4} \frac{1}{2} \sigma_{\max}(\mathcal{T}).$$

Using the same argument as above, one can obtain

$$\left\| \boldsymbol{E}_1^{0,\ell} \right\| \leq \frac{1}{2^{20} \kappa^6 \mu^2 r^4} \frac{1}{2} . \sigma_{\max}(\mathcal{T}).$$

A.1.2 Bounding $\left\| \boldsymbol{X}_i^1 \boldsymbol{R}_i^1 - \boldsymbol{U}_i \right\|_{2,\infty}$

Theorem 3.1 in Cai et al., 2021a shows that with high probability,

$$\begin{aligned}
\left\| \boldsymbol{X}_i^1 \boldsymbol{R}_i^1 - \boldsymbol{U}_i \right\|_{2,\infty} &\leq C\kappa^2 \left(\frac{\kappa^2 \mu^{3/2} r^{3/2} \log n}{n^{3/2} p} + \sqrt{\frac{\kappa^4 \mu^2 r^2 \log n}{n^2 p}} + \frac{\kappa^2 \mu r}{n} \right) \sqrt{\frac{\mu r}{n}} \\
&\leq \frac{1}{2} \frac{1}{2^{20} \kappa^2 \mu^2 r^4} \frac{1}{2} \sqrt{\frac{\mu r}{n}},
\end{aligned}$$

where the last inequality follows from the assumption $n \geq C_0 \kappa^6 \mu^3 r^5$ and

$$p \geq \max \left\{ \frac{C_1 \kappa^6 \mu^{3.5} r^{5.5} \log n}{n^{3/2}}, \frac{C_2 \kappa^{12} \mu^6 r^{10} \log n}{n^2} \right\}. \qquad (30)$$

A.1.3 Bounding $\left\| \boldsymbol{X}_i^1 \boldsymbol{R}_i^1 - \boldsymbol{X}_i^{1,\ell} \boldsymbol{T}_i^{1,\ell} \right\|_{\mathsf{F}}$

Lemma 4 in Cai et al., 2021a, Lemma 8 in Cai et al., 2021b and Lemma 34 demonstrate that with high probability,

$$\begin{aligned}
\left\| \boldsymbol{X}_i^1 \boldsymbol{R}_i^1 - \boldsymbol{X}_i^{1,\ell} \boldsymbol{T}_i^{1,\ell} \right\|_{\mathsf{F}} &\leq \left\| \boldsymbol{X}_i^{1,\ell} \boldsymbol{X}_i^{1,\ell\top} - \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top} \right\|_{\mathsf{F}} \\
&\leq C \left(\frac{\kappa^2 \mu^{3/2} r^{3/2} \log n}{n^{3/2} p} + \sqrt{\frac{\kappa^4 \mu^2 r^2 \log n}{n^2 p}} \right) \sqrt{\frac{\mu r}{n}} \\
&\leq \frac{1}{2^{25} \kappa^2 \mu^2 r^4} \frac{1}{2} \sqrt{\frac{\mu r}{n}}, \qquad (31)
\end{aligned}$$

where the last line is due to the assumption $p \geq \max \left\{ \frac{C_1 \kappa^4 \mu^{3.5} r^{5.5} \log n}{n^{3/2}}, \frac{C_2 \kappa^8 \mu^6 r^{10} \log n}{n^2} \right\}$. The inequality (31) directly implies that (17e) holds for $t = 1$.

A.1.4 Bounding $\left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}-\boldsymbol{U}_i\right\|_{2,\infty}$

Invoking the triangle inequality gives

$$\begin{aligned}\left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}-\boldsymbol{U}_i\right\|_{2,\infty} &\le \left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}-\boldsymbol{X}_i^{1}\boldsymbol{R}_i^{1}\right\|_{2,\infty}+\left\|\boldsymbol{X}_i^{1}\boldsymbol{R}_i^{1}-\boldsymbol{U}_i\right\|_{2,\infty}\\ &\le \left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}-\boldsymbol{X}_i^{1}\boldsymbol{R}_i^{1}\right\|_{\mathsf{F}}+\left\|\boldsymbol{X}_i^{1}\boldsymbol{R}_i^{1}-\boldsymbol{U}_i\right\|_{2,\infty}. \end{aligned}\tag{32}$$

By (26) and (31), the following inequalities hold with high probability,

$$\left\|\boldsymbol{X}_i^{1}\boldsymbol{R}_i^{1}-\boldsymbol{U}_i\right\|\left\|\boldsymbol{U}_i\right\|\le\frac{1}{2}\quad\text{and}\quad\left\|\boldsymbol{X}_i^{1}\boldsymbol{R}_i^{1}-\boldsymbol{X}_i^{1,\ell}\boldsymbol{T}_i^{1,\ell}\right\|\left\|\boldsymbol{U}_i\right\|\le\frac{1}{4}.$$

Applying Lemma 32 with $\boldsymbol{X}_1=\boldsymbol{X}_i^{1}\boldsymbol{R}_i^{1},\ \boldsymbol{X}_2=\boldsymbol{X}_i^{1,\ell}\boldsymbol{T}_i^{1,\ell}$, one can see that

$$\left\|\boldsymbol{X}_i^{1}\boldsymbol{R}_i^{1}-\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}\right\|_{\mathsf{F}}\le 5\left\|\boldsymbol{X}_i^{1}\boldsymbol{R}_i^{1}-\boldsymbol{X}_i^{1,\ell}\boldsymbol{T}_i^{1,\ell}\right\|_{\mathsf{F}}\le\frac{5}{2^{25}\kappa^2\mu^2r^4}\frac{1}{2}\sqrt{\frac{\mu r}{n}},\tag{33}$$

where the last line is due to (31). Plugging (30) and (33) into (32) gives

$$\left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}-\boldsymbol{U}_i\right\|_{2,\infty}\le\frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2}\sqrt{\frac{\mu r}{n}}.\tag{34}$$

A.1.5 Bounding $\left\|\mathcal{E}^0-\mathcal{E}^{0,\ell}\right\|_{\mathsf{F}}$

By the definitions of $\mathcal{E}^0$ and $\mathcal{E}^{0,\ell}$ in (14) and (15), we can decompose $\mathcal{E}^0-\mathcal{E}^{0,\ell}$ as

$$\begin{aligned}\mathcal{E}^0-\mathcal{E}^{0,\ell}&=\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T})\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{1}\boldsymbol{X}_i^{1\top}-\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T})\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell\top}\\ &\quad+\left(\left(p^{-1}\mathcal{P}_\Omega-p^{-1}\mathcal{P}_{\Omega_{-\ell}}-\mathcal{P}_\ell\right)(\mathcal{T})\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell\top}\\ &\quad+\mathcal{T}\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{1}\boldsymbol{X}_i^{1\top}-\mathcal{T}\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell\top}.\end{aligned}$$

It then follows from the triangle inequality that

$$\begin{aligned}\left\|\mathcal{E}^0-\mathcal{E}^{0,\ell}\right\|_{\mathsf{F}}&\le\underbrace{\left\|\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T})\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{1}\boldsymbol{X}_i^{1\top}-\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T})\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell\top}\right\|_{\mathsf{F}}}_{=:\varphi_3}\\ &\quad+\underbrace{\left\|\left(\left(p^{-1}\mathcal{P}_\Omega-p^{-1}\mathcal{P}_{\Omega_{-\ell}}-\mathcal{P}_\ell\right)(\mathcal{T})\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell\top}\right\|_{\mathsf{F}}}_{=:\varphi_4}\\ &\quad+\underbrace{\left\|\mathcal{T}\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{1}\boldsymbol{X}_i^{1\top}-\mathcal{T}\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell\top}\right\|_{\mathsf{F}}}_{=:\varphi_5}.\end{aligned}$$

Bounding φ_3. Simple calculation reveals that

$$
\begin{aligned}
&\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T})\right) \times_{i=1}^{3} \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top} - \left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T})\right) \times_{i=1}^{3} \boldsymbol{X}_i^{1,\ell} \boldsymbol{X}_i^{1,\ell\top} \\
&\quad = \left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T})\right) \times_1 \left(\boldsymbol{X}_1^1 \boldsymbol{X}_1^{1\top} - \boldsymbol{X}_1^{1,\ell} \boldsymbol{X}_1^{1,\ell}\right) \times_2 \boldsymbol{X}_2^{1,\ell} \boldsymbol{X}_2^{1,\ell\top} \times_3 \boldsymbol{X}_3^{1,\ell} \boldsymbol{X}_3^{1,\ell\top} \\
&\qquad + \left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T})\right) \times_1 \boldsymbol{X}_1^1 \boldsymbol{X}_1^{1\top} \times_2 \left(\boldsymbol{X}_2^1 \boldsymbol{X}_2^{1\top} - \boldsymbol{X}_2^{1,\ell} \boldsymbol{X}_2^{1,\ell\top}\right) \times_3 \boldsymbol{X}_3^{1,\ell} \boldsymbol{X}_3^{1,\ell\top} \\
&\qquad + \left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(-\mathcal{T})\right) \times_1 \boldsymbol{X}_1^1 \boldsymbol{X}_1^{1\top} \times_2 \boldsymbol{X}_2^1 \boldsymbol{X}_2^{1\top} \times_3 \left(\boldsymbol{X}_3^1 \boldsymbol{X}_3^{1\top} - \boldsymbol{X}_3^{1,\ell} \boldsymbol{X}_3^{1,\ell\top}\right).
\end{aligned}
$$

Thus we have

$$
\varphi_3 \le \sum_{i=1}^{3} \left\| \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top} - \boldsymbol{X}_i^{1,\ell} \boldsymbol{X}_i^{1,\ell\top} \right\|_{\mathsf{F}} \cdot \sqrt{r} \left\| \left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{T}) \right\|,
$$

where the inequality is due to Lemma 22. Together with (28) and (31), one can see that

$$
\begin{aligned}
\varphi_3 &\le 3 \max_{i=1,2,3} \left\| \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top} - \boldsymbol{X}_i^{1,\ell} \boldsymbol{X}_i^{1,\ell\top} \right\|_{\mathsf{F}} \cdot \sqrt{r} \left\| \left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{T}) \right\| \\
&\le 3 \cdot \frac{1}{2^{25}\kappa^2\mu^2 r^4} \frac{1}{2} \sqrt{\frac{\mu r}{n}} \cdot \sqrt{r} \left(\frac{\mu^{3/2} r^{3/2} \log^3 n}{n^{3/2} p} + \sqrt{\frac{\mu^2 r^2 \log^5 n}{n^2 p}} \right) \sigma_{\max}(\mathcal{T}) \\
&\le \frac{1}{4} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}),
\end{aligned} \tag{35}
$$

where it is assumed that $p \ge \max\left\{ \frac{C_1 \kappa^4 \mu^{3.5} r^{5.5} \log n}{n^{3/2}}, \frac{C_2 \kappa^8 \mu^6 r^{10} \log n}{n^2} \right\}$.

Bounding φ_4. Since $\|\mathcal{Z}\|_{\mathsf{F}} = \|\mathcal{M}_i(\mathcal{Z})\|_{\mathsf{F}}$ for any tensor $\mathcal{Z}$, one has

$$
\begin{aligned}
\varphi_4 &= \left\| \boldsymbol{X}_1^{1,\ell} \boldsymbol{X}_1^{1,\ell\top} \mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_\Omega - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)(\mathcal{T})\right) \left(\boldsymbol{X}_3^{1,\ell} \otimes \boldsymbol{X}_2^{1,\ell}\right) \left(\boldsymbol{X}_3^{1,\ell} \otimes \boldsymbol{X}_2^{1,\ell}\right)^{\top} \right\|_{\mathsf{F}} \\
&\le \left\| \boldsymbol{X}_1^{1,\ell\top} \mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_\Omega - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)(\mathcal{T})\right) \left(\boldsymbol{X}_3^{1,\ell} \otimes \boldsymbol{X}_2^{1,\ell}\right) \right\|_{\mathsf{F}}.
\end{aligned}
$$

By the definition of $\mathcal{P}_{\Omega_{-\ell}}$ and $\mathcal{P}_\ell$, it can be seen that the entries of $\mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_\Omega - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)(\mathcal{T})\right)$ are all zeros except those on the ℓ-th row or the j-th column for any $j \in \Gamma$, where Γ is an index set defined by

$$
\Gamma = \{\ell, n+\ell, \cdots, n(\ell-2)+\ell, n(\ell-1)+1, \cdots, n(\ell-1)+n, n\ell+\ell, \cdots, n(n-1)+\ell\}.
$$

Using this fact, we have

$$
\begin{aligned}
\varphi_4 &\le \left\| \mathcal{P}_{\ell,:}\left(\mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_\Omega - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)(\mathcal{T})\right)\right) \left(\boldsymbol{X}_3^{1,\ell} \otimes \boldsymbol{X}_2^{1,\ell}\right) \right\|_{\mathsf{F}} \\
&\quad + \left\| \boldsymbol{X}_1^{1,\ell\top} \mathcal{P}_{-\ell,\Gamma}\left(\mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_\Omega - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)(\mathcal{T})\right)\right) \left(\boldsymbol{X}_3^{1,\ell} \otimes \boldsymbol{X}_2^{1,\ell}\right) \right\|_{\mathsf{F}} \\
&:= \varphi_{4,1} + \varphi_{4,2}.
\end{aligned}
$$

Here for any matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n^2}$, $\mathcal{P}_{\ell,:}(\boldsymbol{M})$ and $\mathcal{P}_{-\ell,\Gamma}(\boldsymbol{M})$ are defined by

$$[\mathcal{P}_{\ell,:}(\boldsymbol{M})]_{i,j} = \begin{cases} [\boldsymbol{M}]_{i,j}, & \text{if } i = \ell, \\ 0, & \text{otherwise}, \end{cases} \tag{36}$$

$$[\mathcal{P}_{-\ell,\Gamma}(\boldsymbol{M})]_{i,j} = \begin{cases} [\boldsymbol{M}]_{i,j}, & \text{if } i \neq \ell \text{ and } j \in \Gamma, \\ 0, & \text{otherwise}. \end{cases} \tag{37}$$

- For $\varphi_{4,1}$, note that

$$\begin{aligned} &\mathcal{P}_{\ell,:}\left(\mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_\Omega - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)(\mathcal{T})\right)\right)\left(\boldsymbol{X}_3^{1,\ell} \otimes \boldsymbol{X}_2^{1,\ell}\right) \\ &= \sum_{j=1}^{n^2} \left(1 - p^{-1}\delta_{\ell,j}\right)[\mathcal{M}_1(\mathcal{T})]_{\ell,j}\left[\boldsymbol{X}_3^{1,\ell} \otimes \boldsymbol{X}_2^{1,\ell}\right]_{j,:} := \sum_{j=1}^{n^2} \boldsymbol{x}_j^\top, \end{aligned}$$

where $\boldsymbol{X}_2^{1,\ell}$ and $\boldsymbol{X}_3^{1,\ell}$ are independent of $\{\delta_{\ell,j}\}_{j\in[n^2]}$ by construction. A simple computation implies that

$$\begin{aligned} \|\boldsymbol{x}_j\| &\leq p^{-1}\|\mathcal{T}\|_\infty \cdot \left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^{1,\ell}\right\|_{2,\infty}\right)^2, \\ \left\|\sum_{j=1}^{n^2}\mathbb{E}\left\{\boldsymbol{x}_j^\top \boldsymbol{x}_j\right\}\right\| &\leq p^{-1}\|\boldsymbol{T}_1\|_{2,\infty}^2 \cdot \left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^{1,\ell}\right\|_{2,\infty}\right)^4, \\ \left\|\sum_{j=1}^{n^2}\mathbb{E}\left\{\boldsymbol{x}_j \boldsymbol{x}_j^\top\right\}\right\| &\leq p^{-1}\|\boldsymbol{T}_1\|_{2,\infty}^2 \cdot \left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^{1,\ell}\right\|_{2,\infty}\right)^4. \end{aligned}$$

By the matrix Bernstein inequality (Tropp, 2015, Theorem 6.1.1), the following inequality holds with high probability,

$$\varphi_{4,1} \leq C\left(\frac{\log n}{p}\|\mathcal{T}\|_\infty \cdot \left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^{1,\ell}\right\|_{2,\infty}\right)^2 + \sqrt{\frac{\log n}{p}\|\boldsymbol{T}_1\|_{2,\infty}^2 \cdot \left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^{1,\ell}\right\|_{2,\infty}\right)^4}\right)$$

Moreover, from (34), we have

$$\left\|\boldsymbol{X}_i^{1,\ell}\right\|_{2,\infty} \leq \left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell} - \boldsymbol{U}_i\right\|_{2,\infty} + \|\boldsymbol{U}_i\|_{2,\infty} \leq 2\sqrt{\frac{\mu r}{n}}. \tag{38}$$

Therefore, with high probability,

$$\begin{aligned} \varphi_{4,1} &\leq C\left(\frac{4\log n}{p}\left(\frac{\mu r}{n}\right)^{5/2} + \sqrt{\frac{16\log n}{p}\left(\frac{\mu r}{n}\right)^3}\right)\sigma_{\max}(\mathcal{T}) \\ &\leq \frac{1}{8}\frac{1}{2^{20}\kappa^4\mu^2 r^4}\frac{1}{2}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \end{aligned}$$

where the last step is due to the assumption $p \geq \frac{C_2\kappa^8\mu^6 r^{10}\log n}{n^2}$.

- For $\varphi_{4,2}$, it can be rearranged as follows:

$$\begin{aligned}
\varphi_{4,2} &\leq \sqrt{r} \left\| \mathcal{P}_{-\ell,\Gamma} \left(\mathcal{M}_1 \left(\left(p^{-1}\mathcal{P}_{\Omega} - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_{\ell} \right) (\mathcal{T}) \right) \right) \left(\boldsymbol{X}_3^{1,\ell} \otimes \boldsymbol{X}_2^{1,\ell} \right) \right\| \\
&= \sqrt{r} \left\| \sum_{j \in \Gamma} \underbrace{\left[\mathcal{P}_{-\ell,:} \left(\mathcal{M}_1 \left(\left(p^{-1}\mathcal{P}_{\Omega} - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_{\ell} \right) (\mathcal{T}) \right) \right) \right]_{:,j} \left[\boldsymbol{X}_3^{1,\ell} \otimes \boldsymbol{X}_2^{1,\ell} \right]_{j,:}}_{:= \boldsymbol{Z}_j} \right\|,
\end{aligned}$$

where $\boldsymbol{Z}_j$ are independent mean-zero random matrices conditioned on $\boldsymbol{X}_2^{1,\ell}$ and $\boldsymbol{X}_3^{1,\ell}$. First, a straightforward computation shows that

$$\begin{aligned}
\|\boldsymbol{Z}_j\| &\leq \left\| \left[\mathcal{P}_{-\ell,:} \left(\mathcal{M}_1 \left(\left(p^{-1}\mathcal{P}_{\Omega} - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_{\ell} \right) (\mathcal{T}) \right) \right) \right]_{:,j} \right\|_2 \cdot \left\| \left[\boldsymbol{X}_3^{1,\ell} \otimes \boldsymbol{X}_2^{1,\ell} \right]_{j,:} \right\|_2 \\
&\leq \left\| \left[\mathcal{M}_1 \left(\left(p^{-1}\mathcal{P}_{\Omega} - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_{\ell} \right) (\mathcal{T}) \right) \right]_{:,j} \right\|_2 \cdot \left\| \left[\boldsymbol{X}_3^{1,\ell} \otimes \boldsymbol{X}_2^{1,\ell} \right]_{j,:} \right\|_2 \\
&\leq \frac{1}{p} \left\| \boldsymbol{T}_1^{\mathsf{T}} \right\|_{2,\infty} \cdot \left(\max_{i=1,2,3} \left\| \boldsymbol{X}_i^{1,\ell} \right\|_{2,\infty} \right)^2.
\end{aligned}$$

Moreover, one has

$$\begin{aligned}
\sigma_{\boldsymbol{Z}}^2 &:= \left\| \sum_{j \in \Gamma} \mathbb{E} \left\{ \boldsymbol{Z}_j \boldsymbol{Z}_j^{\mathsf{T}} \right\} \right\| \leq \frac{1}{p} \left(\max_{i=1,2,3} \left\| \boldsymbol{X}_i^{1,\ell} \right\|_{2,\infty} \right)^4 \cdot \|\boldsymbol{T}_1\|_{2,\infty}^2 \leq \frac{16}{p} \cdot \left(\frac{\mu r}{n} \right)^3 \sigma_{\max}^2(\mathcal{T}), \\
\sigma_{\boldsymbol{Z}'}^2 &:= \left\| \sum_{j \in \Gamma} \mathbb{E} \left\{ \boldsymbol{Z}_j^{\mathsf{T}} \boldsymbol{Z}_j \right\} \right\| \leq \frac{2n}{p} \left\| \boldsymbol{T}_1^{\mathsf{T}} \right\|_{2,\infty}^2 \cdot \left(\max_{i=1,2,3} \left\| \boldsymbol{X}_i^{1,\ell} \right\|_{2,\infty} \right)^4 \leq \frac{32n}{p} \cdot \left(\frac{\mu r}{n} \right)^4 \sigma_{\max}^2(\mathcal{T}).
\end{aligned}$$

Therefore, by the matrix Bernstein inequality, we know that

$$\begin{aligned}
\varphi_{4,2} &\leq \sqrt{r} \cdot C \left(\frac{\log n}{p} \left\| \boldsymbol{T}_1^{\mathsf{T}} \right\|_{2,\infty} \cdot \left(\max_{i=1,2,3} \left\| \boldsymbol{X}_i^{1,\ell} \right\|_{2,\infty} \right)^2 + \sqrt{\max\left\{ \sigma_{\boldsymbol{Z}}^2, \sigma_{\boldsymbol{Z}'}^2 \right\} \log n} \right) \\
&\leq \sqrt{r} \cdot C \left(\frac{\log n}{p} \left(\frac{\mu r}{n} \right)^{3/2} + \sqrt{\frac{n \log n}{p} \cdot \left(\frac{\mu r}{n} \right)^3} \right) \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}) \\
&\leq \frac{1}{8} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T})
\end{aligned}$$

holds with high probability under the assumption $p \geq \max\left\{ \frac{C_1 \kappa^4 \mu^{3.5} r^6 \log n}{n^{3/2}}, \frac{C_2 \kappa^8 \mu^7 r^{12} \log n}{n^2} \right\}$.

Combining the above bounds together, one can find that

$$\varphi_4 \leq \frac{1}{4} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}). \tag{39}$$

Bounding φ_5. It follows from the triangle inequality that

$$\begin{aligned}\varphi_5 &\leq \left\| \mathcal{T} \times_1 \left(\boldsymbol{X}_1^1 \boldsymbol{X}_1^{1\top} - \boldsymbol{X}_1^{1,\ell} \boldsymbol{X}_1^{1,\ell\top} \right) \times_2 \boldsymbol{X}_2^{1,\ell} \boldsymbol{X}_2^{1,\ell\top} \times_3 \boldsymbol{X}_3^{1,\ell} \boldsymbol{X}_3^{1,\ell\top} \right\|_{\mathsf{F}} \\ &\quad + \left\| \mathcal{T} \times_1 \boldsymbol{X}_1^1 \boldsymbol{X}_1^{1\top} \times_2 \left(\boldsymbol{X}_2^1 \boldsymbol{X}_2^{1\top} - \boldsymbol{X}_2^{1,\ell} \boldsymbol{X}_2^{1,\ell\top} \right) \times_3 \boldsymbol{X}_3^{1,\ell} \boldsymbol{X}_3^{1,\ell\top} \right\|_{\mathsf{F}} \\ &\quad + \left\| \mathcal{T} \times_1 \boldsymbol{X}_1^1 \boldsymbol{X}_1^{1\top} \times_2 \boldsymbol{X}_2^1 \boldsymbol{X}_2^{1\top} \times_3 \left(\boldsymbol{X}_3^1 \boldsymbol{X}_3^{1\top} - \boldsymbol{X}_3^{1,\ell} \boldsymbol{X}_3^{1,\ell\top} \right) \right\|_{\mathsf{F}} \\ &\leq 3 \max_{i=1,2,3} \left\| \boldsymbol{X}_i^1 \boldsymbol{X}_i^{1\top} - \boldsymbol{X}_i^{1,\ell} \boldsymbol{X}_i^{1,\ell\top} \right\|_{\mathsf{F}} \cdot \sigma_{\max}(\mathcal{T}) \overset{(a)}{\leq} \frac{1}{4} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}), \end{aligned} \tag{40}$$

where (a) follows from (31).

Combining φ_3, φ_4 and φ_5. Putting (35), (39) and (40) together yields that

$$\left\| \mathcal{E}^0 - \mathcal{E}^{0,\ell} \right\|_{\mathsf{F}} \leq \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}).$$

A.2 Induction Step for Theorem 9

For the sake of clarity, a proof roadmap is presented in Figure 2, which shows the dependencies of different quantities during the induction process. We first summarize below several

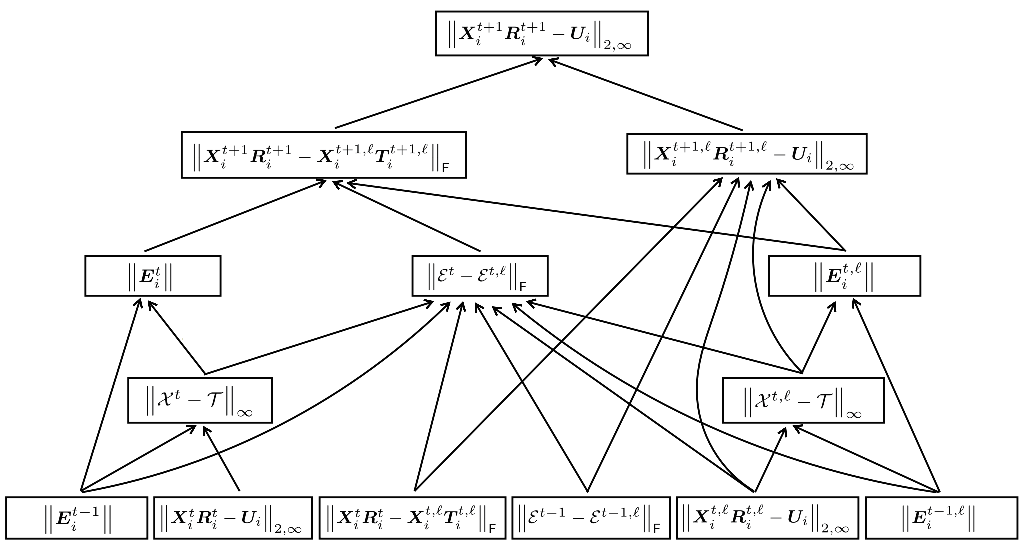

Figure 2: Proof roadmap for the induction step.

immediate consequences of (17), which will be useful throughout this section. The proofs of these results are provided in Section B.

Lemma 18 *From the inequalities* (17a), (17b) *and* (17c), *one has*

$$\left\| \boldsymbol{X}_i^t \boldsymbol{R}_i^t - \boldsymbol{U}_i \right\| \leq \frac{1}{2^{17}\kappa^4\mu^2 r^4} \frac{1}{2^t}, \quad i = 1, 2, 3, \tag{41}$$

$$\left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{R}_i^{t,\ell}-\boldsymbol{U}_i\right\| \leq \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2^t}, \quad i=1,2,3, \tag{42}$$

$$\left\|\boldsymbol{X}_i^{t}\boldsymbol{X}_i^{t\mathsf{T}}-\boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\right\| \leq \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2^t}, \quad i=1,2,3, \tag{43}$$

$$\left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\mathsf{T}}-\boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\right\| \leq \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2^t}, \quad i=1,2,3, \tag{44}$$

$$\left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\mathsf{T}}-\boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\right\|_{2,\infty} \leq \frac{1}{2^{16}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}, \quad i=1,2,3. \tag{45}$$

Lemma 19 *From the inequalities* (17a), (17b), (17c) *and* (17f), *one has*

$$\left\|\mathcal{X}^t-\mathcal{T}\right\|_\infty \leq \frac{36}{2^{20}\kappa^2\mu^2r^4}\cdot\frac{1}{2^t}\cdot\left(\frac{\mu r}{n}\right)^{3/2}\cdot\sigma_{\max}(\mathcal{T}), \tag{46}$$

$$\left\|\mathcal{X}^{t,\ell}-\mathcal{T}\right\|_\infty \leq \frac{36}{2^{20}\kappa^2\mu^2r^4}\cdot\frac{1}{2^t}\cdot\left(\frac{\mu r}{n}\right)^{3/2}\cdot\sigma_{\max}(\mathcal{T}). \tag{47}$$

Lemma 20 *Assuming* (17a) *and* (17b), *the largest and the smallest nonzero singular values of* $\mathcal{M}_i\left(\mathcal{X}^t\right)$ *satisfy*

$$\left(1+\frac{1}{2^9}\right)\sigma_{\max}(\mathcal{T}) \geq \sigma_{\max}\left(\mathcal{M}_i\left(\mathcal{X}^t\right)\right) \geq \left(1-\frac{1}{2^9}\right)\sigma_{\min}(\mathcal{T}) \geq \frac{15}{16}\sigma_{\min}(\mathcal{T}), \tag{48}$$

$$\left(1+\frac{1}{2^9}\right)\sigma_{\max}(\mathcal{T}) \geq \sigma_{\max}\left(\mathcal{M}_i\left(\mathcal{X}^{t,\ell}\right)\right) \geq \left(1-\frac{1}{2^9}\right)\sigma_{\min}(\mathcal{T}) \geq \frac{15}{16}\sigma_{\min}(\mathcal{T}). \tag{49}$$

It follows immediately that the condition number of $\mathcal{X}^t$ *obeys*

$$\kappa\left(\mathcal{X}^t\right) = \frac{\sigma_{\max}\left(\mathcal{X}^t\right)}{\sigma_{\min}\left(\mathcal{X}^t\right)} \leq \frac{\left(1+2^{-9}\right)\sigma_{\max}\left(\mathcal{T}\right)}{\left(1-2^{-9}\right)\sigma_{\min}\left(\mathcal{T}\right)} \leq 2\kappa.$$

A.2.1 Bounding $\left\|\boldsymbol{E}_i^t\right\|$ and $\left\|\boldsymbol{E}_i^{t,\ell}\right\|$

We provide a detailed proof for $\left\|\boldsymbol{E}_1^t\right\|$, while the proofs for $\left\|\boldsymbol{E}_i^t\right\|$ $(i=2,3)$ and $\left\|\boldsymbol{E}_i^{t,\ell}\right\|$ $(i=1,2,3)$ are overall similar. First, $\left\|\boldsymbol{E}_1^t\right\|$ can be bounded as follows:

$$\begin{aligned}
\left\|\boldsymbol{E}_1^t\right\| &\leq \left\|\mathcal{M}_1\left(\left(\mathcal{I}-\mathcal{P}_{T_{\mathcal{X}^t}}\right)\left(\mathcal{X}^t-\mathcal{T}\right)\right)\right\| + \left\|\mathcal{M}_1\left(\mathcal{P}_{T_{\mathcal{X}^t}}\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t-\mathcal{T}\right)\right)\right\| \\
&\leq \left\|\mathcal{M}_1\left(\left(\mathcal{I}-\mathcal{P}_{T_{\mathcal{X}^t}}\right)\left(\mathcal{X}^t-\mathcal{T}\right)\right)\right\|_{\mathsf{F}} + \left\|\mathcal{M}_1\left(\mathcal{P}_{T_{\mathcal{X}^t}}\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t-\mathcal{T}\right)\right)\right\| \\
&= \left\|\left(\mathcal{I}-\mathcal{P}_{T_{\mathcal{X}^t}}\right)\left(\mathcal{X}^t-\mathcal{T}\right)\right\|_{\mathsf{F}} + \left\|\mathcal{M}_1\left(\mathcal{P}_{T_{\mathcal{X}^t}}\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t-\mathcal{T}\right)\right)\right\| \\
&= \underbrace{\left\|\left(\mathcal{I}-\mathcal{P}_{T_{\mathcal{X}^t}}\right)\left(\mathcal{T}\right)\right\|_{\mathsf{F}}}_{=:\alpha_1} + \underbrace{\left\|\mathcal{M}_1\left(\mathcal{P}_{T_{\mathcal{X}^t}}\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t-\mathcal{T}\right)\right)\right\|}_{=:\alpha_2}.
\end{aligned}$$

To proceed, we need to decompose $\mathcal{P}_{T_{\mathcal{X}^t}}$ into a sum of products of several projectors (Cai et al., 2020). Define the projector $\mathcal{P}_{\boldsymbol{X}_i^t}^{(i)}:\mathbb{R}^{n\times n\times n}\rightarrow\mathbb{R}^{n\times n\times n}$ by

$$\mathcal{P}_{\boldsymbol{X}_i^t}^{(i)}(\mathcal{Z}) = \mathcal{Z}\times_i\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}},\ \forall\mathcal{Z}\in\mathbb{R}^{n\times n\times n}. \tag{50}$$

It is evident that $\mathcal{Z} \mathop{\times}\limits_{i=1}^{3} \boldsymbol{X}_i^t {\boldsymbol{X}_i^t}^\top = \prod_{i=1}^3 \mathcal{P}_{\boldsymbol{X}_i^t}^{(i)}(\mathcal{Z})$. The orthogonal complement of $\mathcal{P}_{\boldsymbol{X}_i^t}^{(i)}$, denoted $\mathcal{P}_{\boldsymbol{X}_i^{t,\perp}}^{(i)} : \mathbb{R}^{n\times n\times n} \to \mathbb{R}^{n\times n\times n}$, is defined as

$$\mathcal{P}_{\boldsymbol{X}_i^{t,\perp}}^{(i)}(\mathcal{Z}) = \mathcal{Z} \times_i \left(\boldsymbol{I} - \boldsymbol{X}_i^t {\boldsymbol{X}_i^t}^\top\right).$$

Furthermore, we define $\mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq 1}}^{(j\neq 1)} : \mathbb{R}^{n\times n\times n} \to \mathbb{R}^{n\times n\times n}$ by

$$\mathcal{M}_1\left(\mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq 1}}^{(j\neq 1)}(\mathcal{Z})\right) = \mathcal{M}_1(\mathcal{Z})\left(\boldsymbol{X}_3^t \otimes \boldsymbol{X}_2^t\right)\mathcal{M}_1^\dagger(\mathcal{G}^t)\mathcal{M}_1(\mathcal{G}^t)\left(\boldsymbol{X}_3^t \otimes \boldsymbol{X}_2^t\right)^\top, \tag{51}$$

and $\mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq 2}}^{(j\neq 2)}$, $\mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq 3}}^{(j\neq 3)}$ are similarly defined. All together, it is not hard to see that $\mathcal{P}_{T_{\mathcal{X}^t}}$ can be rewritten as

$$\mathcal{P}_{T_{\mathcal{X}^t}} = \prod_{i=1}^3 \mathcal{P}_{\boldsymbol{X}_i^t}^{(i)} + \sum_{i=1}^3 \mathcal{P}_{\boldsymbol{X}_i^{t,\perp}}^{(i)} \mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)}. \tag{52}$$

Moreover, the following properties hold, whose proofs are provided in Section C.1.

Claim A.1 *By the definition of $\mathcal{P}_{\boldsymbol{X}_i^t}^{(i)}$, $\mathcal{P}_{\boldsymbol{X}_i^{t,\perp}}^{(i)}$ and $\mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)}$, one has*

$$\mathcal{P}_{\boldsymbol{X}_i^{t,\perp}}^{(i)} \mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)} = \mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)} \mathcal{P}_{\boldsymbol{X}_i^{t,\perp}}^{(i)}, \quad \mathcal{P}_{\boldsymbol{X}_i^t}^{(i)} \mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)} = \mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)} \mathcal{P}_{\boldsymbol{X}_i^t}^{(i)},$$

$$\mathcal{M}_i\left(\mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)}(\mathcal{Z})\right) = \mathcal{M}_i(\mathcal{Z})\,\boldsymbol{Y}_i^t {\boldsymbol{Y}_i^t}^\top, \quad \left\|\prod_{j\neq i} \mathcal{P}_{\boldsymbol{X}_j^t}^{(j)} - \mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)}\right\| = 1,$$

$$\left(\prod_{j\neq i} \mathcal{P}_{\boldsymbol{X}_j^t}^{(j)} - \mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)}\right)(\mathcal{X}^t) = \boldsymbol{0},$$

where the columns of $\boldsymbol{Y}_i^t$ are the top-r right singular vectors of $\mathcal{M}_i\left(\mathcal{X}^t\right)$.

Bounding α_1. Based on the decomposition of $\mathcal{P}_{T_{\mathcal{X}^t}}$, it has been pointed out in Cai et al., 2020 that

$$\begin{aligned}
\mathcal{I} - \mathcal{P}_{T_{\mathcal{X}^t}} &= \prod_{i=1}^3\left(\mathcal{P}_{\boldsymbol{X}_i^t}^{(i)} + \mathcal{P}_{\boldsymbol{X}_i^{t,\perp}}^{(i)}\right) - \left(\prod_{i=1}^3 \mathcal{P}_{\boldsymbol{X}_i^t}^{(i)} + \sum_{i=1}^3 \mathcal{P}_{\boldsymbol{X}_i^{t,\perp}}^{(i)} \mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)}\right) \\
&= \sum_{i=1}^3 \mathcal{P}_{\boldsymbol{X}_i^{t,\perp}}^{(i)}\left(\prod_{j\neq i} \mathcal{P}_{\boldsymbol{X}_j^t}^{(j)} - \mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)}\right) + \sum_{i=1}^3 \mathcal{P}_{\boldsymbol{X}_i^t}^{(i)} \prod_{j\neq i} \mathcal{P}_{\boldsymbol{X}_j^{t,\perp}}^{(j)} + \prod_{i=1}^3 \mathcal{P}_{\boldsymbol{X}_i^{t,\perp}}^{(i)}.
\end{aligned}$$

Note that, for any $i = 1,2,3$,

$$\mathcal{P}_{\boldsymbol{X}_i^{t,\perp}}^{(i)}\left(\mathcal{X}^t\right) = \boldsymbol{0} \quad \text{and} \quad \left(\prod_{j\neq i} \mathcal{P}_{\boldsymbol{X}_j^t}^{(j)} - \mathcal{P}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}}^{(j\neq i)}\right)\left(\mathcal{X}^t\right) = \boldsymbol{0}.$$

Therefore,

$$\begin{aligned}
\left(\mathcal{I}-\mathcal{P}_{T_{\mathcal{X}^t}}\right)(\mathcal{T}) &= \sum_{i=1}^{3}\left(\mathcal{P}_{\boldsymbol{U}_i}^{(i)}-\mathcal{P}_{\boldsymbol{X}_i^t}^{(i)}\right)\left(\prod_{j\neq i}\mathcal{P}_{\boldsymbol{X}_j^t}^{(j)}-\mathcal{P}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq i}}^{(j\neq i)}\right)(\mathcal{T}) \\
&\quad+\left(\mathcal{P}_{\boldsymbol{U}_1}^{(1)}-\mathcal{P}_{\boldsymbol{X}_1^t}^{(1)}\right)\mathcal{P}_{\boldsymbol{X}_2^{t,\perp}}^{(2)}\mathcal{P}_{\boldsymbol{X}_3^t}^{(3)}(\mathcal{T})+\left(\mathcal{P}_{\boldsymbol{U}_2}^{(2)}-\mathcal{P}_{\boldsymbol{X}_2^t}^{(2)}\right)\mathcal{P}_{\boldsymbol{X}_1^t}^{(1)}\mathcal{P}_{\boldsymbol{X}_3^{t,\perp}}^{(3)}(\mathcal{T}) \\
&\quad+\left(\mathcal{P}_{\boldsymbol{U}_3}^{(3)}-\mathcal{P}_{\boldsymbol{X}_3^t}^{(3)}\right)\mathcal{P}_{\boldsymbol{X}_1^{t,\perp}}^{(1)}\mathcal{P}_{\boldsymbol{X}_2^t}^{(2)}(\mathcal{T})+\left(\mathcal{P}_{\boldsymbol{U}_1}^{(1)}-\mathcal{P}_{\boldsymbol{X}_1^t}^{(1)}\right)\mathcal{P}_{\boldsymbol{X}_2^{t,\perp}}^{(2)}\mathcal{P}_{\boldsymbol{X}_3^{t,\perp}}^{(3)}(\mathcal{T}) \\
&= -\sum_{i=1}^{3}\left(\mathcal{P}_{\boldsymbol{U}_i}^{(i)}-\mathcal{P}_{\boldsymbol{X}_i^t}^{(i)}\right)\left(\prod_{j\neq i}\mathcal{P}_{\boldsymbol{X}_j^t}^{(j)}-\mathcal{P}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq i}}^{(j\neq i)}\right)(\mathcal{X}^t-\mathcal{T}) \\
&\quad-\left(\mathcal{P}_{\boldsymbol{U}_1}^{(1)}-\mathcal{P}_{\boldsymbol{X}_1^t}^{(1)}\right)\mathcal{P}_{\boldsymbol{X}_2^{t,\perp}}^{(2)}\mathcal{P}_{\boldsymbol{X}_3^t}^{(3)}(\mathcal{X}^t-\mathcal{T})-\left(\mathcal{P}_{\boldsymbol{U}_2}^{(2)}-\mathcal{P}_{\boldsymbol{X}_2^t}^{(2)}\right)\mathcal{P}_{\boldsymbol{X}_1^t}^{(1)}\mathcal{P}_{\boldsymbol{X}_3^{t,\perp}}^{(3)}(\mathcal{X}^t-\mathcal{T}) \\
&\quad-\left(\mathcal{P}_{\boldsymbol{U}_3}^{(3)}-\mathcal{P}_{\boldsymbol{X}_3^t}^{(3)}\right)\mathcal{P}_{\boldsymbol{X}_1^{t,\perp}}^{(1)}\mathcal{P}_{\boldsymbol{X}_2^t}^{(2)}(\mathcal{X}^t-\mathcal{T})-\left(\mathcal{P}_{\boldsymbol{U}_1}^{(1)}-\mathcal{P}_{\boldsymbol{X}_1^t}^{(1)}\right)\mathcal{P}_{\boldsymbol{X}_2^{t,\perp}}^{(2)}\mathcal{P}_{\boldsymbol{X}_3^{t,\perp}}^{(3)}(\mathcal{X}^t-\mathcal{T}).
\end{aligned} \tag{53}$$

It follows that

$$\begin{aligned}
\alpha_1 &\leq \sum_{i=1}^{3}\left\|\boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}-\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}}\right\|\cdot\left\|\prod_{j\neq i}\mathcal{P}_{\boldsymbol{X}_j^t}^{(j)}-\mathcal{P}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq i}}^{(j\neq i)}\right\|\cdot\left\|\mathcal{X}^t-\mathcal{T}\right\|_{\mathsf{F}} \\
&\quad+4\max_{i=1,2,3}\left\|\boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}-\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}}\right\|\cdot\left\|\mathcal{X}^t-\mathcal{T}\right\|_{\mathsf{F}} \\
&\overset{(a)}{\leq} 7\max_{i=1,2,3}\left\|\boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}-\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}}\right\|\cdot\left\|\mathcal{X}^t-\mathcal{T}\right\|_{\mathsf{F}} \\
&\leq 7\max_{i=1,2,3}\left\|\boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}-\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}}\right\|\cdot n^{3/2}\cdot\left\|\mathcal{X}^t-\mathcal{T}\right\|_{\infty} \\
&\overset{(b)}{\leq} 7\cdot\frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2^t}\cdot n^{3/2}\cdot\frac{36}{2^{20}\kappa^2\mu^2r^4}\cdot\frac{1}{2^t}\cdot\left(\frac{\mu r}{n}\right)^{3/2}\cdot\sigma_{\max}(\mathcal{T}) \\
&\leq \frac{1}{2^9}\frac{1}{2^{20}\kappa^6\mu^2r^4}\cdot\frac{1}{2^{t+1}}\cdot\sigma_{\max}(\mathcal{T}),
\end{aligned} \tag{54}$$

where (a) is due to Claim A.1 and (b) follows from (43) and (46).

Bounding α_2. By the definition of $\mathcal{P}_{T_{\mathcal{X}^t}}$ in (8) and the triangular inequality, we have

$$\alpha_2 \leq \underbrace{\left\|\mathcal{M}_1\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{X}^t-\mathcal{T})\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}}\right)\right\|}_{=:\alpha_{2,1}}+\underbrace{\sum_{i=1}^{3}\left\|\mathcal{M}_1\left(\mathcal{G}^t\times_i\boldsymbol{W}_i^t\mathop{\times}_{j\neq i}\boldsymbol{X}_j^t\right)\right\|}_{=:\alpha_{2,2}},$$

where $\boldsymbol{W}_i^t$ is given by

$$\boldsymbol{W}_i^t=\left(\boldsymbol{I}-\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}}\right)\mathcal{M}_i\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{X}^t-\mathcal{T})\mathop{\times}_{j\neq i}\boldsymbol{X}_j^{t\mathsf{T}}\right)\mathcal{M}_i^{\dagger}(\mathcal{G}^t),\quad i=1,2,3.$$

- Controlling $\alpha_{2,1}$. Notice that the tensor $\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t-\mathcal{T}\right)\times_{i=1}^{3}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top}$ has multilinear rank at most (r,r,r). Applying Lemma 22 yields that

$$\begin{aligned}\alpha_{2,1} &\le \sqrt{r}\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t-\mathcal{T}\right)\times_{i=1}^{3}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top}\right\| \le \sqrt{r}\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t-\mathcal{T}\right)\right\|\\ &\overset{(a)}{\le} \left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right\|\cdot 2r^{3/2}\left\|\mathcal{X}^t-\mathcal{T}\right\|_\infty,\end{aligned}$$

where (a) follows from Lemma 25 and the fact that $\mathcal{X}^t-\mathcal{T}$ has multilinear rank at most $(2r,2r,2r)$. Here $\mathcal{J}$ is the all-one tensor.

- Controlling $\alpha_{2,2}$. Straightforward computation gives that

$$\begin{aligned}\alpha_{2,2} &\le \left\|\mathcal{M}_1(\mathcal{G}^t)\right\|\cdot\left(\sum_{i=1}^{3}\left\|\boldsymbol{W}_i^t\right\|\right)\\ &\le \left\|\mathcal{M}_1(\mathcal{G}^t)\right\|\cdot\left(\sum_{i=1}^{3}\left\|\mathcal{M}_i\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t-\mathcal{T}\right)\times_{j\neq i}\boldsymbol{X}_j^{t\top}\right)\right\|\cdot\left\|\mathcal{M}_i^{\dagger}(\mathcal{G}^t)\right\|\right)\\ &\overset{(a)}{\le} \left\|\mathcal{M}_1(\mathcal{G}^t)\right\|\cdot\left(\sum_{i=1}^{3}\sqrt{r}\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t-\mathcal{T}\right)\right\|\cdot\left\|\mathcal{M}_i^{\dagger}(\mathcal{G}^t)\right\|\right)\\ &\le 3\left\|\mathcal{M}_1(\mathcal{G}^t)\right\|\cdot\sqrt{r}\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t-\mathcal{T}\right)\right\|\cdot\max_{i=1,2,3}\left\|\mathcal{M}_i^{\dagger}(\mathcal{G}^t)\right\|\\ &\le 3\sqrt{r}\cdot\frac{\max_{i=1,2,3}\sigma_{\max}\left(\mathcal{M}_i(\mathcal{G}^t)\right)}{\min_{i=1,2,3}\sigma_{\min}\left(\mathcal{M}_i(\mathcal{G}^t)\right)}\cdot\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t-\mathcal{T}\right)\right\|\\ &\le 6r^{3/2}\cdot\kappa\left(\mathcal{X}^t\right)\cdot\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right\|\cdot\left\|\mathcal{X}^t-\mathcal{T}\right\|_\infty\\ &\overset{(b)}{\le} 6r^{3/2}\cdot 2\kappa\cdot\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right\|\cdot\left\|\mathcal{X}^t-\mathcal{T}\right\|_\infty,\end{aligned}$$

where (a) is due to Lemma 22 and (b) follows from Lemma 20.

Combining the upper bounds of $\alpha_{2,1}$ and $\alpha_{2,2}$ together yields that, with high probability,

$$\begin{aligned}\alpha_2 &\le \alpha_{2,1}+\alpha_{2,2} \le 14r^{3/2}\kappa\cdot\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right\|\cdot\left\|\mathcal{X}^t-\mathcal{T}\right\|_\infty\\ &\overset{(a)}{\le} 14r^{3/2}\kappa\cdot C\left(\frac{\log^3 n}{p}+\sqrt{\frac{n\log^5 n}{p}}\right)\cdot\frac{36}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\left(\frac{\mu r}{n}\right)^{3/2}\sigma_{\max}(\mathcal{T})\\ &\le \frac{1}{2^9}\frac{1}{2^{20}\kappa^6\mu^2r^4}\cdot\frac{1}{2^{t+1}}\cdot\sigma_{\max}(\mathcal{T}),\end{aligned}$$

where (a) follows from Lemma 23 and the assumption $p\ge\max\left\{\frac{C_1\kappa^5\mu^{3/2}r^3\log^3 n}{n^{3/2}},\frac{C_2\kappa^{10}\mu^3r^6\log^5 n}{n^2}\right\}$.

Putting the upper bounds of α_1 and α_2 together, one has

$$\left\|\boldsymbol{E}_1^t\right\| \le \frac{1}{2^8}\frac{1}{2^{20}\kappa^6\mu^2r^4}\cdot\frac{1}{2^{t+1}}\cdot\sigma_{\max}(\mathcal{T}) \le \frac{1}{2^{20}\kappa^6\mu^2r^4}\cdot\frac{1}{2^{t+1}}\cdot\sigma_{\max}(\mathcal{T}) \le \frac{1}{4}\sigma_{\max}(\mathcal{T}). \tag{55}$$

Using the same argument as above, one can obtain

$$\left\|\boldsymbol{E}_1^{t,\ell}\right\| \le \frac{1}{2^8}\frac{1}{2^{20}\kappa^6\mu^2r^4}\cdot\frac{1}{2^{t+1}}\cdot\sigma_{\max}(\mathcal{T}) \le \frac{1}{2^{20}\kappa^6\mu^2r^4}\cdot\frac{1}{2^{t+1}}\cdot\sigma_{\max}(\mathcal{T}) \le \frac{1}{4}\sigma_{\max}(\mathcal{T}). \tag{56}$$

A.2.2 Bounding $\left\|\boldsymbol{X}_i^{t+1,\ell}\boldsymbol{R}_i^{t+1,\ell}-\boldsymbol{U}_i\right\|_{2,\infty}$

We only provide a detailed proof for the case $i=1$, and the proofs for the other two cases are overall similar. Since $\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{\Sigma}_1^{t+1,\ell}\boldsymbol{X}_1^{t+1,\ell^\top}$ is the top-r eigenvalue decomposition of

$$\left(\boldsymbol{T}_1+\boldsymbol{E}_1^{t,\ell}\right)\left(\boldsymbol{T}_1+\boldsymbol{E}_1^{t,\ell}\right)^\top=\boldsymbol{T}_1\boldsymbol{T}_1^\top+\underbrace{\boldsymbol{T}_1\boldsymbol{E}_1^{t,\ell^\top}+\boldsymbol{E}_1^{t,\ell}\boldsymbol{T}_1^\top+\boldsymbol{E}_1^{t,\ell}\boldsymbol{E}_1^{t,\ell^\top}}_{=:\boldsymbol{\Delta}_1^{t,\ell}},$$

we have

$$\boldsymbol{X}_1^{t+1,\ell}=\left(\boldsymbol{T}_1\boldsymbol{T}_1^\top+\boldsymbol{\Delta}_1^{t,\ell}\right)\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}.$$

Recall that the eigenvalue decomposition of $\boldsymbol{T}_1\boldsymbol{T}_1^\top$ is $\boldsymbol{T}_1\boldsymbol{T}_1^\top=\boldsymbol{U}_1\boldsymbol{\Lambda}_1\boldsymbol{U}_1^\top$, $\boldsymbol{e}_m^\top\left(\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{R}_1^{t+1,\ell}-\boldsymbol{U}_1\right)$ can be decomposed as

$$\begin{aligned}
&\boldsymbol{e}_m^\top\left(\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{R}_1^{t+1,\ell}-\boldsymbol{U}_1\right)\\
&=\boldsymbol{e}_m^\top\left(\boldsymbol{T}_1\boldsymbol{T}_1^\top\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\boldsymbol{R}_1^{t+1,\ell}-\boldsymbol{U}_1\right)+\boldsymbol{e}_m^\top\boldsymbol{\Delta}_1^{t,\ell}\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\boldsymbol{R}_1^{t+1,\ell}\\
&=\boldsymbol{e}_m^\top\boldsymbol{U}_1\boldsymbol{\Lambda}_1\left(\boldsymbol{U}_1^\top\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\boldsymbol{R}_1^{t+1,\ell}-\boldsymbol{\Lambda}_1^{-1}\right)+\boldsymbol{e}_m^\top\boldsymbol{\Delta}_1^{t,\ell}\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\boldsymbol{R}_1^{t+1,\ell}\\
&=\boldsymbol{e}_m^\top\boldsymbol{U}_1\boldsymbol{\Lambda}_1\left(\boldsymbol{U}_1^\top\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{\Lambda}_1^{-1}\boldsymbol{R}_1^{t+1,\ell}-\boldsymbol{\Lambda}_1^{-1}\right)\\
&\quad+\boldsymbol{e}_m^\top\boldsymbol{U}_1\boldsymbol{\Lambda}_1\boldsymbol{U}_1^\top\boldsymbol{X}_1^{t+1,\ell}\left(\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}-\boldsymbol{\Lambda}_1^{-1}\right)\boldsymbol{R}_1^{t+1,\ell}+\boldsymbol{e}_m^\top\boldsymbol{\Delta}_1^{t,\ell}\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\boldsymbol{R}_1^{t+1,\ell}\\
&=\boldsymbol{e}_m^\top\boldsymbol{U}_1\left(\boldsymbol{\Lambda}_1\boldsymbol{U}_1^\top\boldsymbol{X}_1^{t+1,\ell}-\boldsymbol{R}_1^{t+1,\ell^\top}\boldsymbol{\Lambda}_1^{-1}\right)\boldsymbol{\Lambda}_1^{-1}\boldsymbol{R}_1^{t+1,\ell}\\
&\quad+\boldsymbol{e}_m^\top\boldsymbol{U}_1\boldsymbol{\Lambda}_1\boldsymbol{U}_1^\top\boldsymbol{X}_1^{t+1,\ell}\left(\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}-\boldsymbol{\Lambda}_1^{-1}\right)\boldsymbol{R}_1^{t+1,\ell}+\boldsymbol{e}_m^\top\boldsymbol{\Delta}_1^{t,\ell}\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\boldsymbol{R}_1^{t+1,\ell}.
\end{aligned}$$

The triangle inequality then gives that

$$\begin{aligned}
\left\|\boldsymbol{e}_m^\top\left(\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{R}_1^{t+1,\ell}-\boldsymbol{U}_1\right)\right\|_2\le&\underbrace{\left\|\boldsymbol{e}_m^\top\boldsymbol{U}_1\left(\boldsymbol{\Lambda}_1\boldsymbol{U}_1^\top\boldsymbol{X}_1^{t+1,\ell}-\boldsymbol{R}_1^{t+1,\ell^\top}\boldsymbol{\Lambda}_1\right)\boldsymbol{\Lambda}_1^{-1}\boldsymbol{R}_1^{t+1,\ell}\right\|_2}_{=:\beta_1}\\
&+\underbrace{\left\|\boldsymbol{e}_m^\top\boldsymbol{U}_1\boldsymbol{\Lambda}_1\boldsymbol{U}_1^\top\boldsymbol{X}_1^{t+1,\ell}\left(\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}-\boldsymbol{\Lambda}_1^{-1}\right)\boldsymbol{R}_1^{t+1,\ell}\right\|_2}_{=:\beta_2}\\
&+\underbrace{\left\|\boldsymbol{e}_m^\top\boldsymbol{\Delta}_1^{t,\ell}\boldsymbol{X}_1^{t+1,\ell}\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\boldsymbol{R}_1^{t+1,\ell}\right\|_2}_{=:\beta_3}.
\end{aligned}\tag{57}$$

Bounding β_1. Notice that, with high probability,

$$\begin{aligned}
\left\|\boldsymbol{\Delta}_1^{t,\ell}\right\|=\left\|\boldsymbol{T}_1\boldsymbol{E}_1^{t,\ell^\top}+\boldsymbol{E}_1^{t,\ell}\boldsymbol{T}_1^\top+\boldsymbol{E}_1^{t,\ell}\boldsymbol{E}_1^{t,\ell^\top}\right\|&\le\left(2\left\|\boldsymbol{T}_1\right\|+\left\|\boldsymbol{E}_1^{t,\ell}\right\|\right)\left\|\boldsymbol{E}_1^{t,\ell}\right\|\\
&\overset{(a)}{\le}\frac{9}{4}\sigma_{\max}(\mathcal{T})\cdot\frac{1}{2^8}\frac{1}{2^{20}\kappa^6\mu^2r^4}\cdot\frac{1}{2^{t+1}}\cdot\sigma_{\max}(\mathcal{T})
\end{aligned}$$

$$= \frac{9}{2^{10}} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2^{t+1}} \sigma_{\min}^2(\mathcal{T}) \le \frac{1}{2}\sigma_{\min}^2(\mathcal{T}), \tag{58}$$

where step (a) follows from (56). Given the SVD of $\boldsymbol{H}_1^{t+1,\ell} := \boldsymbol{X}_1^{t+1,\ell\top}\boldsymbol{U}_1$, denoted $\boldsymbol{H}_1^{t+1,\ell} = \widehat{\boldsymbol{A}}_1^{t+1,\ell}\widehat{\boldsymbol{\Sigma}}_1^{t+1,\ell}\widehat{\boldsymbol{B}}_1^{t+1,\ell\top}$, it can be easily shown that $\boldsymbol{R}_1^{t+1,\ell} = \widehat{\boldsymbol{A}}_1^{t+1,\ell}\widehat{\boldsymbol{B}}_1^{t+1,\ell\top}$. Therefore, the application of Lemma 36 yields that

$$\begin{aligned}
&\left\|\left(\boldsymbol{\Lambda}_1\boldsymbol{U}_1^\top\boldsymbol{X}_1^{t+1,\ell} - \boldsymbol{R}_1^{t+1,\ell\top}\boldsymbol{\Lambda}_1\right)\boldsymbol{\Lambda}_1^{-1}\right\| \\
&\quad \le \left\|\boldsymbol{\Lambda}_1\boldsymbol{R}_1^{t+1,\ell} - \boldsymbol{H}_1^{t+1,\ell}\boldsymbol{\Lambda}_1\right\| \left\|\boldsymbol{\Lambda}_1^{-1}\right\| \le \left(2 + \frac{\sigma_{\max}^2(\mathcal{T})}{\sigma_{\min}^2(\mathcal{T}) - \left\|\boldsymbol{\Delta}_1^{t,\ell}\right\|}\right)\left\|\boldsymbol{\Delta}_1^{t,\ell}\right\| \frac{1}{\sigma_{\min}\left(\boldsymbol{\Lambda}_1\right)} \\
&\quad \le \left(2 + 2\kappa^2\right)\cdot \frac{9}{2^{10}} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2^{t+1}} \le \frac{36}{2^{10}} \frac{1}{2^{20}\kappa^2\mu^2 r^4} \frac{1}{2^{t+1}},
\end{aligned} \tag{59}$$

which occurs with high probability. Here the third line is due to the fact $\sigma_{\min}(\boldsymbol{\Lambda}_1) \ge \sigma_{\min}^2(\mathcal{T})$ and (58). Thus β_1 can be bounded with high probability as follows:

$$\beta_1 \le \|\boldsymbol{U}_1\|_{2,\infty}\left\|\left(\boldsymbol{\Lambda}_1\boldsymbol{U}_1^\top\boldsymbol{X}_1^{t+1,\ell} - \boldsymbol{R}_1^{t+1,\ell\top}\boldsymbol{\Lambda}_1\right)\boldsymbol{\Lambda}_1^{-1}\right\| \le \frac{36}{2^{10}} \frac{1}{2^{20}\kappa^2\mu^2 r^4} \frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}. \tag{60}$$

Bounding β_2. Applying the Weyl's inequality yields that $\left\|\boldsymbol{\Lambda}_1 - \boldsymbol{\Sigma}_1^{t+1,\ell}\right\| \le \left\|\boldsymbol{\Delta}_1^{t,\ell}\right\|$. It follows that

$$\sigma_r\left(\boldsymbol{\Sigma}_1^{t+1,\ell}\right) \ge \sigma_{\min}^2(\mathcal{T}) - \left\|\boldsymbol{\Delta}_1^{t,\ell}\right\| \ge \frac{1}{2}\sigma_{\min}^2(\mathcal{T}),$$

and

$$\begin{aligned}
\left\|\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}} - \boldsymbol{\Lambda}_1^{-1}\right\| &= \left\|\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\left(\boldsymbol{\Lambda}_1 - \boldsymbol{\Sigma}_1^{t+1,\ell}\right)\boldsymbol{\Lambda}_1^{-1}\right\| \le \left\|\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\right\|\left\|\boldsymbol{\Lambda}_1^{-1}\right\|\left\|\boldsymbol{\Lambda}_1 - \boldsymbol{\Sigma}_1^{t+1,\ell}\right\| \\
&\le \frac{1}{\sigma_r\left(\boldsymbol{\Sigma}_1^{t+1,\ell}\right)}\cdot\frac{1}{\sigma_{\min}^2(\mathcal{T})}\cdot\left\|\boldsymbol{\Lambda}_1 - \boldsymbol{\Sigma}_1^{t+1,\ell}\right\| \le \frac{2}{\sigma_{\min}^4(\mathcal{T})}\cdot\left\|\boldsymbol{\Delta}_1^{t,\ell}\right\| \\
&\le \frac{2}{\sigma_{\min}^4(\mathcal{T})}\cdot\frac{9}{2^{10}} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2^{t+1}}\sigma_{\min}^2(\mathcal{T}) = \frac{2}{\sigma_{\min}^2\left(\mathcal{T}\right)}\frac{9}{2^{10}} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2^{t+1}}.
\end{aligned}$$

Therefore, with high probability,

$$\beta_2 \le \|\boldsymbol{U}_1\|_{2,\infty}\cdot\|\boldsymbol{\Lambda}_1\|\cdot\left\|\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}} - \boldsymbol{\Lambda}_1^{-1}\right\| \le \frac{18}{2^{10}} \frac{1}{2^{20}\kappa^2\mu^2 r^4} \frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}. \tag{61}$$

Bounding β_3. The last term β_3 can be bounded as follows:

$$\begin{aligned}
\beta_3 &\le \left\|\boldsymbol{e}_m^\top\boldsymbol{\Delta}_1^{t,\ell}\right\|_2\cdot\left\|\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\right\| = \left\|\boldsymbol{e}_m^\top\left(\boldsymbol{T}_1\boldsymbol{E}_1^{t,\ell\top} + \boldsymbol{E}_1^{t,\ell}\boldsymbol{T}_1^\top + \boldsymbol{E}_1^{t,\ell}\boldsymbol{E}_1^{t,\ell\top}\right)\right\|_{2,\infty}\cdot\left\|\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\right\| \\
&\le \left(\|\boldsymbol{T}_1\|_{2,\infty}\cdot\left\|\boldsymbol{E}_1^{t,\ell}\right\| + \left\|\boldsymbol{e}_m^\top\boldsymbol{E}_1^{t,\ell}\right\|_2\cdot\left(\|\boldsymbol{T}_1\| + \left\|\boldsymbol{E}_1^{t,\ell}\right\|\right)\right)\cdot\left\|\boldsymbol{\Sigma}_1^{t+1,\ell^{-1}}\right\| \\
&\le \sqrt{\frac{\mu r}{n}}\cdot\frac{1}{2^8}\frac{1}{2^{20}\kappa^6\mu^2 r^4}\frac{1}{2^{t+1}}\sigma_{\max}^2(\mathcal{T})\cdot\frac{2}{\sigma_{\min}^2(\mathcal{T})} + \left\|\boldsymbol{e}_m^\top\boldsymbol{E}_1^{t,\ell}\right\|_2\cdot\frac{3\sigma_{\max}(\mathcal{T})}{\sigma_{\min}^2(\mathcal{T})}
\end{aligned}$$

$$\leq \frac{1}{2^7} \frac{1}{2^{20}\kappa^2\mu^2 r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} + \left\| \boldsymbol{e}_m^\top \boldsymbol{E}_1^{t,\ell} \right\|_2 \cdot \frac{3\sigma_{\max}(\mathcal{T})}{\sigma_{\min}^2(\mathcal{T})}, \tag{62}$$

where the third inequality is due to (56) and

$$\left(\|\boldsymbol{T}_1\| + \left\| \boldsymbol{E}_1^{t,\ell} \right\| \right) \cdot \left\| \boldsymbol{\Sigma}_1^{t+1,\ell-1} \right\| \leq \frac{2}{\sigma_{\min}^2(\mathcal{T})} \cdot \left(\sigma_{\max}(\mathcal{T}) + \frac{1}{4}\sigma_{\max}(\mathcal{T}) \right) \leq \frac{3\sigma_{\max}(\mathcal{T})}{\sigma_{\min}^2(\mathcal{T})}.$$

To this end, we focus on bounding $\left\| \boldsymbol{e}_m^\top \boldsymbol{E}_1^{t,\ell} \right\|_2$. Recall that $\boldsymbol{E}_1^{t,\ell}$ is defined by

$$\boldsymbol{E}_1^{t,\ell} = \mathcal{M}_1 \left(\left(\mathcal{I} - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}} \left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell \right) \right) \left(\mathcal{X}^t - \mathcal{T} \right) \right).$$

Thus, invoking the triangle inequality yields that

$$\begin{aligned} \left\| \boldsymbol{e}_m^\top \boldsymbol{E}_1^{t,\ell} \right\|_2 &\leq \left\| \boldsymbol{e}_m^\top \mathcal{M}_1 \left(\left(\mathcal{I} - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2 \\ &\quad + \left\| \boldsymbol{e}_m^\top \mathcal{M}_1 \left(\mathcal{P}_{T_{\mathcal{X}^{t,\ell}}} \left(\mathcal{I} - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2 =: \beta_{3,a} + \beta_{3,b}. \end{aligned}$$

The following two claims provide the upper bounds for $\beta_{3,a}$ and $\beta_{3,b}$. The proofs of the claims can be found in Section C.2 and Section C.3, respectively.

Claim A.2 *Assuming the inequalities in* (17) *hold, one can obtain*

$$\beta_{3,a} \leq \frac{1}{2^6} \cdot \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}).$$

Claim A.3 *Suppose* $p \geq \frac{C_1 \kappa^8 \mu^{3.5} r^6 \log^3 n}{n^{3/2}}$. *Assuming the inequalities in* (17) *hold, one has*

$$\beta_{3,b} \leq \frac{1}{2^6} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}).$$

Combining Claim A.2 and Claim A.3 together reveals that

$$\left\| \boldsymbol{e}_m^\top \boldsymbol{E}_1^{t,\ell} \right\|_2 \leq \frac{1}{2^5} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}). \tag{63}$$

Furthermore, putting (62) and (63) together yields

$$\beta_3 \leq \frac{1}{2^3} \frac{1}{2^{20}\kappa^2\mu^2 r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}}. \tag{64}$$

Combining β_1, β_2 and β_3. Plugging (60), (61) and (64) into (57) shows that with high probability,

$$\left\| \boldsymbol{e}_m^\top \left(\boldsymbol{X}_1^{t+1,\ell} \boldsymbol{R}_1^{t+1,\ell} - \boldsymbol{U}_1 \right) \right\|_2 \leq \frac{1}{4} \frac{1}{2^{20}\kappa^2\mu^2 r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}}. \tag{65}$$

Taking the maximum of (65) over m gives that

$$\left\| \boldsymbol{X}_1^{t+1,\ell} \boldsymbol{R}_1^{t+1,\ell} - \boldsymbol{U}_1 \right\|_{2,\infty} \leq \frac{1}{4} \frac{1}{2^{20}\kappa^2\mu^2 r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}}.$$

The same bound can be obtained for the other two modes. Thus, we have

$$\left\| \boldsymbol{X}_i^{t+1,\ell} \boldsymbol{R}_i^{t+1,\ell} - \boldsymbol{U}_i \right\|_{2,\infty} \leq \frac{1}{4} \frac{1}{2^{20}\kappa^2\mu^2 r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}}, \quad i = 1, 2, 3. \tag{66}$$

A.2.3 Bounding $\left\|\mathcal{E}^t - \mathcal{E}^{t,\ell}\right\|_{\mathsf{F}}$

By the definition of $\mathcal{E}^t$ and $\mathcal{E}^{t,\ell}$ in (13) and (16),

$$\begin{aligned}
\mathcal{E}^t - \mathcal{E}^{t,\ell} &= \left(\mathcal{I} - p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t - \mathcal{T}\right) - \left(\mathcal{I} - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell\right)\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right) \\
&= \left(\mathcal{I} - \mathcal{P}_{T_{\mathcal{X}^t}}\right)\left(\mathcal{X}^t - \mathcal{T}\right) + \mathcal{P}_{T_{\mathcal{X}^t}}\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t - \mathcal{T}\right) \\
&\quad - \left(\mathcal{I} - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right) + \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{I} - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right) \\
&= \left(\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\right)(\mathcal{T}) + \left(\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\right)\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t - \mathcal{T}\right) \\
&\quad - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t - \mathcal{T}\right) + \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{I} - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right).
\end{aligned}$$

Thus we can bound $\left\|\mathcal{E}^t - \mathcal{E}^{t,\ell}\right\|_{\mathsf{F}}$ as follows:

$$\begin{aligned}
\left\|\mathcal{E}^t - \mathcal{E}^{t,\ell}\right\|_{\mathsf{F}} \le& \underbrace{\left\|\left(\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\right)(\mathcal{T})\right\|_{\mathsf{F}}}_{=:\xi_1} + \underbrace{\left\|\left(\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\right)\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t - \mathcal{T}\right)\right\|_{\mathsf{F}}}_{=:\xi_2} \\
&+ \underbrace{\left\|\mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t - \mathcal{T}\right) - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{I} - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right\|_{\mathsf{F}}}_{=:\xi_3}.
\end{aligned} \tag{67}$$

For the operator $\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}$, one can invoke the definition (52) to deduce that

$$\begin{aligned}
&\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}} \\
&= \left(\prod_{i=1}^3 \mathcal{P}^{(i)}_{\boldsymbol{X}_i^t} - \prod_{i=1}^3 \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}\right) + \left(\sum_{i=1}^3 \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\perp}}\mathcal{P}^{(j\neq i)}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq i}} - \sum_{i=1}^3 \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell,\perp}}\mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}}\right) \\
&= \prod_{i=1}^3\left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^t} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}\right) + \sum_{i=1}^3 \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}\prod_{j\neq i}\left(\mathcal{P}^{(j)}_{\boldsymbol{X}_j^t} - \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}}\right) + \sum_{i=1}^3\left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^t} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}\right)\prod_{j\neq i}\mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} \\
&\quad + \sum_{i=1}^3 \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\perp}}\left(\mathcal{P}^{(j\neq i)}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq i}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}}\right) + \sum_{i=1}^3\left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^t}\right)\mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}} \\
&= \prod_{i=1}^3\left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^t} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}\right) + \sum_{i=1}^3 \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}\prod_{j\neq i}\left(\mathcal{P}^{(j)}_{\boldsymbol{X}_j^t} - \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}}\right) \\
&\quad + \sum_{i=1}^3\left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^t} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}\right)\left(\prod_{j\neq i}\mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}}\right) \\
&\quad + \sum_{i=1}^3 \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\perp}}\left(\mathcal{P}^{(j\neq i)}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq i}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}}\right).
\end{aligned} \tag{68}$$

Bounding ξ_1. By (68), this term can be bounded as follows:

$$\xi_1 \le \underbrace{\left\| \prod_{i=1}^{3} \left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^t} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}} \right)(\mathcal{T}) \right\|_{\mathsf{F}}}_{=:\xi_{1,1}} + \underbrace{\sum_{i=1}^{3} \left\| \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}} \prod_{j\neq i} \left(\mathcal{P}^{(j)}_{\boldsymbol{X}_j^t} - \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} \right)(\mathcal{T}) \right\|_{\mathsf{F}}}_{=:\xi_{1,2}}$$
$$+ \underbrace{\sum_{i=1}^{3} \left\| \left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^t} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}} \right) \left(\prod_{j\neq i} \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell}, \left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}} \right)(\mathcal{T}) \right\|_{\mathsf{F}}}_{=:\xi_{1,3}}$$
$$+ \underbrace{\sum_{i=1}^{3} \left\| \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\perp}} \left(\mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t}, \left\{\boldsymbol{X}_j^{t}\right\}_{j\neq i}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell}, \left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}} \right)(\mathcal{T}) \right\|_{\mathsf{F}}}_{=:\xi_{1,4}}.$$

- Bounding $\xi_{1,1}, \xi_{1,2}$ and $\xi_{1,3}$. By the definition of $\mathcal{P}^{(i)}_{\boldsymbol{X}_i^t}$ and $\mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}$, we have

$$\begin{aligned}
\xi_{1,1} &\le \prod_{i=1}^{3} \left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} \right\|_{\mathsf{F}} \cdot \sigma_{\max}(\mathcal{T}) \\
&\le \left(\max_{i=1,2,3} \left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^{t,\ell} \boldsymbol{X}_i^{t,\ell\top} \right\|_{\mathsf{F}} \right)^3 \cdot \sigma_{\max}(\mathcal{T}), \\
\xi_{1,2} &\le \sum_{i=1}^{3} \left\| \mathcal{M}_i \left(\prod_{j\neq i} \left(\mathcal{P}^{(j)}_{\boldsymbol{X}_j^t} - \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} \right)(\mathcal{T}) \right) \right\|_{\mathsf{F}} \\
&\le \sum_{i=1}^{3} \prod_{j\neq i} \left\| \boldsymbol{X}_j^t \boldsymbol{X}_j^{t\top} - \boldsymbol{X}_j^t \boldsymbol{X}_j^{t\top} \right\|_{\mathsf{F}} \cdot \sigma_{\max}(\mathcal{T}) \\
&\le 3 \left(\max_{i=1,2,3} \left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^{t,\ell} \boldsymbol{X}_i^{t,\ell\top} \right\|_{\mathsf{F}} \right)^2 \cdot \sigma_{\max}(\mathcal{T}), \\
\xi_{1,3} &\overset{(a)}{=} \sum_{i=1}^{3} \left\| \left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^t} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}} \right) \left(\prod_{j\neq i} \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell}, \left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right\|_{\mathsf{F}} \\
&= \sum_{i=1}^{3} \left\| \mathcal{M}_i \left(\left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^t} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}} \right) \left(\prod_{j\neq i} \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell}, \left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_{\mathsf{F}} \\
&\le \sum_{i=1}^{3} \left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^{t,\ell} \boldsymbol{X}_i^{t,\ell\top} \right\|_{\mathsf{F}} \cdot \left\| \left(\prod_{j\neq i} \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell}, \left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right\|_{\mathsf{F}} \\
&\overset{(b)}{\le} \sum_{i=1}^{3} \left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^{t,\ell} \boldsymbol{X}_i^{t,\ell\top} \right\|_{\mathsf{F}} \cdot \left\| \mathcal{X}^{t,\ell} - \mathcal{T} \right\|_{\mathsf{F}} \\
&\le 3 \max_{i=1,2,3} \left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^{t,\ell} \boldsymbol{X}_i^{t,\ell\top} \right\|_{\mathsf{F}} \cdot \left\| \mathcal{X}^{t,\ell} - \mathcal{T} \right\|_{\mathsf{F}},
\end{aligned}$$

where step (a) and step (b) follow from the Claim A.1. Combining these three terms together and using (47), one can see that

$$\begin{aligned}\sum_{i=1}^{3}\xi_{1,i} &\leq \left(\frac{1}{2^{19}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\right)^3\sigma_{\max}(\mathcal{T}) + 3\left(\frac{1}{2^{19}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\right)^2\sigma_{\max}(\mathcal{T}) \\ &\quad + 3\cdot\frac{1}{2^{19}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\cdot n^{3/2}\frac{36}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\left(\frac{\mu r}{n}\right)^{3/2}\sigma_{\max}(\mathcal{T}) \\ &\leq \left(\frac{1}{2^{19}}+\frac{3}{2^{18}}+\frac{3\cdot 36}{2^{19}}\right)\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}) \\ &\leq \frac{1}{2^{10}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \end{aligned} \tag{69}$$

where we have used the following bound

$$\begin{aligned}\left\|\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}} - \boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\mathsf{T}}\right\|_{\mathsf{F}} &= \left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t\left(\boldsymbol{X}_i^t\boldsymbol{R}_i^t\right)^{\mathsf{T}} - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\left(\boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right)^{\mathsf{T}}\right\|_{\mathsf{F}} \\ &\leq 2\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}} \leq 2\cdot\frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}.\end{aligned}$$

- For $\xi_{1,4}$, it follows that

$$\xi_{1,4} = \sum_{i=1}^{3}\left\|\mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\perp}}\left(\mathcal{P}^{(j\neq i)}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell},\{\boldsymbol{X}_j^{t,\ell}\}_{j\neq i}}\right)(\mathcal{T})\right\|_{\mathsf{F}} =: \sum_{i=1}^{3}\xi_{1,4}^i.$$

The proofs for different i are overall similar, so we only provide details for $\xi_{1,4}^1$. The proof starts with the following bound

$$\begin{aligned}\xi_{1,4}^1 &= \left\|\mathcal{M}_1\left(\mathcal{P}^{(1)}_{\boldsymbol{X}_1^{t,\perp}}\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq 1}} - \mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t,\ell},\{\boldsymbol{X}_j^{t,\ell}\}_{j\neq 1}}\right)(\mathcal{T})\right)\right\|_{\mathsf{F}} \\ &= \left\|\left(\boldsymbol{U}_1\boldsymbol{U}_1^{\mathsf{T}} - \boldsymbol{X}_1^t\boldsymbol{X}_1^{t\mathsf{T}}\right)\mathcal{M}_1\left(\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq 1}} - \mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t,\ell},\{\boldsymbol{X}_j^{t,\ell}\}_{j\neq 1}}\right)(\mathcal{T})\right)\right\|_{\mathsf{F}} \\ &\leq \left\|\boldsymbol{U}_1\boldsymbol{U}_1^{\mathsf{T}} - \boldsymbol{X}_1^t\boldsymbol{X}_1^{t\mathsf{T}}\right\|\cdot\left\|\mathcal{M}_1\left(\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq 1}} - \mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t,\ell},\{\boldsymbol{X}_j^{t,\ell}\}_{j\neq 1}}\right)(\mathcal{T})\right)\right\|_{\mathsf{F}} \\ &\leq 2\left\|\boldsymbol{X}_1^t\boldsymbol{R}_1^t - \boldsymbol{U}_1\right\|\cdot\left\|\mathcal{M}_1\left(\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq 1}} - \mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t,\ell},\{\boldsymbol{X}_j^{t,\ell}\}_{j\neq 1}}\right)(\mathcal{T})\right)\right\|_{\mathsf{F}}.\end{aligned} \tag{70}$$

Applying Claim A.1 yields that

$$\begin{aligned}&\left\|\mathcal{M}_1\left(\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq 1}} - \mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t,\ell},\{\boldsymbol{X}_j^{t,\ell}\}_{j\neq 1}}\right)(\mathcal{T})\right)\right\|_{\mathsf{F}} \\ &= \left\|\mathcal{M}_1(\mathcal{T})\left(\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\mathsf{T}} - \boldsymbol{Y}_1^{t,\ell}\boldsymbol{Y}_1^{t,\ell\mathsf{T}}\right)\right\|_{\mathsf{F}} \leq \left\|\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\mathsf{T}} - \boldsymbol{Y}_1^{t,\ell}\boldsymbol{Y}_1^{t,\ell\mathsf{T}}\right\|_{\mathsf{F}}\cdot\sigma_{\max}(\mathcal{T})\end{aligned}$$

$$\overset{(a)}{\leq} \frac{1}{2^{13}} \cdot \frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \tag{71}$$

where the columns of $\boldsymbol{Y}_1^t$ and $\boldsymbol{Y}_1^{t,\ell}$ are the top-r right singular vectors of $\mathcal{M}_1\left(\mathcal{X}^t\right)$ and $\mathcal{M}_1\left(\mathcal{X}^{t,\ell}\right)$, respectively, (a) is due to the following claim, whose proof is presented in Section C.4.

Claim A.4 *Assuming the inequalities in* (17) *hold, one has*

$$\left\|\boldsymbol{Y}_i^t\boldsymbol{Y}_i^{t\mathsf{T}} - \boldsymbol{Y}_i^{t,\ell}\boldsymbol{Y}_i^{t,\ell\mathsf{T}}\right\|_{\mathsf{F}} \leq \frac{1}{2^{13}} \cdot \frac{1}{2^t}\sqrt{\frac{\mu r}{n}}, \quad i = 1,2,3.$$

Plugging (71) into (70) reveals that

$$\xi_{1,4}^1 \leq 2 \cdot \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2^t} \cdot \frac{1}{2^{13}} \cdot \frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}) \leq \frac{1}{2^9} \cdot \frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}).$$

The upper bounds of $\xi_{1,4}^2$ and $\xi_{1,4}^3$ follow immediately via nearly the same argument as above,

$$\xi_{1,4}^2 \leq \frac{1}{2^9} \cdot \frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}) \text{ and } \xi_{1,4}^3 \leq \frac{1}{2^9} \cdot \frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}).$$

Thus, we have

$$\xi_{1,4} \leq \frac{3}{2^9} \cdot \frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}). \tag{72}$$

Combining (69) and (72) together shows that

$$\xi_1 \leq \sum_{i=1}^4 \xi_{1,i} \leq \frac{4}{2^9}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}). \tag{73}$$

Bounding ξ_2. For notational convenience, define

$$\mathcal{Z}^t = \left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t - \mathcal{T}\right).$$

First, apply (68) to deduce that

$$\begin{aligned}
\xi_2 \leq\; & \underbrace{\left\|\prod_{i=1}^3\left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^t} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}\right)\mathcal{Z}^t\right\|_{\mathsf{F}}}_{=:\xi_{2,1}} + \underbrace{\sum_{i=1}^3\left\|\mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}\prod_{j\neq i}\left(\mathcal{P}^{(j)}_{\boldsymbol{X}_j^t} - \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}}\right)\mathcal{Z}^t\right\|_{\mathsf{F}}}_{=:\xi_{2,2}} \\
& + \underbrace{\sum_{i=1}^3\left\|\left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^t} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}\right)\left(\prod_{j\neq i}\mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}}\right)\mathcal{Z}^t\right\|_{\mathsf{F}}}_{=:\xi_{2,3}} \\
& + \underbrace{\sum_{i=1}^3\left\|\mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\perp}}\left(\mathcal{P}^{(j\neq i)}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq i}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}}\right)\mathcal{Z}^t\right\|_{\mathsf{F}}}_{=:\xi_{2,4}}.
\end{aligned} \tag{74}$$

- For $\xi_{2,1}$, it is easily seen that

$$
\begin{aligned}
\xi_{2,1} &\le \prod_{i=1}^{3} \left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^{t,\ell} \boldsymbol{X}_i^{t,\ell\top} \right\|_{\mathsf{F}} \left\| \mathcal{M}_1\left(\mathcal{Z}^t\right) \right\| \overset{(a)}{\le} \prod_{i=1}^{3} \left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^{t,\ell} \boldsymbol{X}_i^{t,\ell\top} \right\|_{\mathsf{F}} \sqrt{n} \left\| \mathcal{Z}^t \right\| \\
&\le 8 \left(\max_{i=1,2,3} \left\| \boldsymbol{X}_i^t \boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell} \boldsymbol{T}_i^{t,\ell} \right\|_{\mathsf{F}} \right)^3 \cdot \sqrt{n} \left\| \mathcal{Z}^t \right\|, \qquad (75)
\end{aligned}
$$

  where (a) follows from Lemma 22.

- For $\xi_{2,2}$, it can be bounded by the same argument as above,

$$
\begin{aligned}
\xi_{2,2} &= \sum_{i=1}^{3} \left\| \mathcal{M}_1 \left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}} \prod_{j \ne i} \left(\mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t}} - \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} \right) \mathcal{Z}^t \right) \right\|_{\mathsf{F}} \\
&\le \sum_{i=1}^{3} \prod_{j \ne i} \left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^{t,\ell} \boldsymbol{X}_i^{t,\ell\top} \right\|_{\mathsf{F}} \cdot \left\| \mathcal{M}_1\left(\mathcal{Z}^t\right) \right\| \\
&\le 3 \cdot 4 \left(\max_{i=1,2,3} \left\| \boldsymbol{X}_i^t \boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell} \boldsymbol{T}_i^{t,\ell} \right\|_{\mathsf{F}} \right)^2 \cdot \sqrt{n} \left\| \mathcal{Z}^t \right\|, \qquad (76)
\end{aligned}
$$

- For $\xi_{2,3}$, letting

$$
\mathcal{Z}_i = \left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t}} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}} \right) \left(\prod_{j \ne i} \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} - \mathcal{P}^{(j \ne i)}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j \ne i}} \right) \mathcal{Z}^t,
$$

  then we have $\xi_{2,3} = \sum_{i=1}^{3} \left\| \mathcal{Z}_i \right\|_{\mathsf{F}}$. It is not hard to see that $\mathcal{Z}_i$ is a tensor of multilinear rank at most $(2r, 2r, 2r)$ for $i = 1, 2, 3$. Thus, the application of Lemma 21 yields that

$$
\begin{aligned}
\xi_{2,3} &\le 4 r^{3/2} \sum_{i=1}^{3} \left\| \left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t}} - \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}} \right) \left(\prod_{j \ne i} \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} - \mathcal{P}^{(j \ne i)}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j \ne i}} \right) \left(\mathcal{Z}^t\right) \right\| \\
&\le 4 r^{3/2} \sum_{i=1}^{3} \left\| \left(\prod_{j \ne i} \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} - \mathcal{P}^{(j \ne i)}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j \ne i}} \right) \left(\mathcal{Z}^t\right) \right\| \cdot \left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^{t,\ell} \boldsymbol{X}_i^{t,\ell\top} \right\|.
\end{aligned}
$$

  To this end, we focus on bounding

$$
\xi_{2,3}^{i} := \left\| \left(\prod_{j \ne i} \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} - \mathcal{P}^{(j \ne i)}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j \ne i}} \right) \left(\mathcal{Z}^t\right) \right\|
$$

  for $i = 1, 2, 3$. The triangular inequality yields that

$$
\xi_{2,3}^{1} \le \left\| \left(\prod_{j \ne 1} \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} \right) \left(\mathcal{Z}^t\right) \right\| + \left\| \mathcal{P}^{(j \ne 1)}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j \ne 1}} \left(\mathcal{Z}^t\right) \right\| := \xi_{2,3}^{1,a} + \xi_{2,3}^{1,b}.
$$

– Bounding $\xi_{2,3}^{1,a}$. It can be bounded as follows:

$$\xi_{2,3}^{1,a} \leq \left\|\mathcal{Z}^t\right\| \cdot \prod_{j\neq 1}\left\|\boldsymbol{X}_j^{t,\ell}\boldsymbol{X}_j^{t,\ell\mathsf{T}}\right\| \leq \left\|\mathcal{Z}^t\right\|. \tag{77}$$

– Bounding $\xi_{2,3}^{1,b}$. A simple computation yields that

$$\begin{aligned}
\xi_{2,3}^{1,b} &\leq \left\|\mathcal{M}_1\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq 1}}\left(\mathcal{Z}^t\right)\right)\right\| \\
&\overset{(a)}{=} \left\|\mathcal{M}_1\left(\mathcal{Z}^t\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)\mathcal{M}_1^{\dagger}\left(\mathcal{G}^{t,\ell}\right)\mathcal{M}_1\left(\mathcal{G}^{t,\ell}\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)^{\mathsf{T}}\right\| \\
&\leq \left\|\mathcal{M}_1\left(\mathcal{Z}^t\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)\right\|\cdot\left\|\mathcal{M}_1^{\dagger}\left(\mathcal{G}^{t,\ell}\right)\mathcal{M}_1\left(\mathcal{G}^{t,\ell}\right)\right\|\cdot\left\|\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right\| \\
&\leq \left\|\mathcal{M}_1\left(\mathcal{Z}^t\times_2\boldsymbol{X}_2^{t,\ell\mathsf{T}}\times_3\boldsymbol{X}_3^{t,\ell\mathsf{T}}\right)\right\| \\
&\overset{(b)}{\leq} \sqrt{r}\left\|\mathcal{Z}^t\times_2\boldsymbol{X}_2^{t,\ell\mathsf{T}}\times_3\boldsymbol{X}_3^{t,\ell\mathsf{T}}\right\| \leq \sqrt{r}\left\|\mathcal{Z}^t\right\|,
\end{aligned} \tag{78}$$

where (a) is due to the definition of $\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq 1}}$ and (b) follows from Lemma 22.

Combining (77) and (78) together yields that

$$\xi_{2,3}^{1} \leq \left\|\mathcal{Z}^t\right\| + \sqrt{r}\left\|\mathcal{Z}^t\right\| \leq 2\sqrt{r}\left\|\mathcal{Z}^t\right\|.$$

Applying the same argument as above can show that

$$\xi_{2,3}^{i} \leq 2\sqrt{r}\left\|\mathcal{Z}^t\right\|, \quad i = 2, 3.$$

Thus one has

$$\begin{aligned}
\xi_{2,3} &\leq 4r^{3/2}\cdot 3\cdot 2\sqrt{r}\left\|\mathcal{Z}^t\right\|\cdot\max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}} - \boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\mathsf{T}}\right\| \\
&= 24r^2\left\|\mathcal{Z}^t\right\|\cdot\max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}} - \boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\mathsf{T}}\right\|_{\mathsf{F}}.
\end{aligned} \tag{79}$$

- For $\xi_{2,4}$, it is easy to see that

$$\begin{aligned}
\xi_{2,4} &= \sum_{i=1}^{3}\left\|\mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\perp}}\left(\mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t},\left\{\boldsymbol{X}_j^{t}\right\}_{j\neq i}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}}\right)\left(\mathcal{Z}^t\right)\right\|_{\mathsf{F}} \\
&\leq \sum_{i=1}^{3}\left\|\mathcal{M}_i\left(\left(\mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t},\left\{\boldsymbol{X}_j^{t}\right\}_{j\neq i}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}}\right)\left(\mathcal{Z}^t\right)\right)\right\|_{\mathsf{F}}.
\end{aligned} \tag{80}$$

The first term in (80) can be bounded as follows:

$$\left\|\mathcal{M}_1\left(\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t},\left\{\boldsymbol{X}_j^{t}\right\}_{j\neq 1}} - \mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t,\ell},\left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq 1}}\right)\left(\mathcal{Z}^t\right)\right)\right\|_{\mathsf{F}}$$

$$\overset{(a)}{\leq} \left\|\mathcal{M}_1(\mathcal{Z}^t)\left(\left(\boldsymbol{X}_3^t\boldsymbol{X}_3^{t\top}\otimes\boldsymbol{X}_2^t\boldsymbol{X}_2^{t\top}\right)\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\top}-\left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\otimes\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\right)\boldsymbol{Y}_1^{t,\ell}\boldsymbol{Y}_1^{t,\ell\top}\right)\right\|_{\mathsf{F}}$$
$$\leq \left\|\mathcal{M}_1(\mathcal{Z}^t)\left(\boldsymbol{X}_3^t\boldsymbol{X}_3^{t\top}\otimes\boldsymbol{X}_2^t\boldsymbol{X}_2^{t\top}-\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\otimes\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\right)\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\top}\right\|_{\mathsf{F}}$$
$$+\left\|\mathcal{M}_1(\mathcal{Z}^t)\left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\otimes\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\right)\left(\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\top}-\boldsymbol{Y}_1^{t,\ell}\boldsymbol{Y}_1^{t,\ell\top}\right)\right\|_{\mathsf{F}} =: \xi_{2,4}^a+\xi_{2,4}^b,$$

where (a) is due to (112).

- Controlling $\xi_{2,4}^a$. Notice that the matrix

$$\mathcal{M}_1(\mathcal{Z}^t)\left(\boldsymbol{X}_3^t\boldsymbol{X}_3^{t\top}\otimes\boldsymbol{X}_2^t\boldsymbol{X}_2^{t\top}-\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\otimes\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\right)\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\top}\in\mathbb{R}^{n\times n^2}$$

is of rank at most r. Letting

$$\boldsymbol{D}_i^t=\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top}-\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top},$$

then $\xi_{2,4}^a$ can be bounded as follows:

$$\xi_{2,4}^a\leq\sqrt{r}\left\|\mathcal{M}_1(\mathcal{Z}^t)\left(\boldsymbol{X}_3^t\boldsymbol{X}_3^{t\top}\otimes\boldsymbol{X}_2^t\boldsymbol{X}_2^{t\top}-\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\otimes\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\right)\right\|$$
$$\leq\sqrt{r}\left\|\mathcal{M}_1(\mathcal{Z}^t)\left(\boldsymbol{D}_3^t\otimes\boldsymbol{X}_2^t\boldsymbol{X}_2^{t\top}\right)\right\|+\sqrt{r}\left\|\mathcal{M}_1(\mathcal{Z}^t)\left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\otimes\boldsymbol{D}_2^t\right)\right\|$$
$$=\sqrt{r}\left\|\mathcal{M}_1\left(\mathcal{Z}^t\times_2\boldsymbol{X}_2^t\boldsymbol{X}_2^{t\top}\times_3\boldsymbol{D}_3^t\right)\right\|+\sqrt{r}\left\|\mathcal{M}_1\left(\mathcal{Z}^t\times_2\boldsymbol{D}_2^t\times_3\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\right)\right\|$$
$$\overset{(a)}{\leq} r\left\|\mathcal{Z}^t\times_2\boldsymbol{X}_2^t\boldsymbol{X}_2^{t\top}\times_3\boldsymbol{D}_3^t\right\|+r\left\|\mathcal{Z}^t\times_2\boldsymbol{D}_2^t\times_3\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\right\|$$
$$\leq 2r\max_{i=2,3}\left\|\boldsymbol{D}_i^t\right\|_{\mathsf{F}}\cdot\left\|\mathcal{Z}^t\right\|\leq 4r\max_{i=2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t-\boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}}\cdot\left\|\mathcal{Z}^t\right\|,$$

where (a) follows from Lemma 22.

- Controlling $\xi_{2,4}^b$. A simple calculation yields that

$$\xi_{2,4}^b\leq 2\sqrt{r}\left\|\mathcal{M}_1(\mathcal{Z}^t)\left(\left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\right)\otimes\left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\right)\right)\left(\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\top}-\boldsymbol{Y}_1^{t,\ell}\boldsymbol{Y}_1^{t,\ell\top}\right)\right\|$$
$$\leq 2\sqrt{r}\left\|\mathcal{M}_1(\mathcal{Z}^t)\left(\left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\right)\otimes\left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\right)\right)\right\|\cdot\left\|\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\top}-\boldsymbol{Y}_1^{t,\ell}\boldsymbol{Y}_1^{t,\ell\top}\right\|$$
$$\leq 2r\left\|\mathcal{Z}^t\times_2\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\times_3\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\right\|\cdot\left\|\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\top}-\boldsymbol{Y}_1^{t,\ell}\boldsymbol{Y}_1^{t,\ell\top}\right\|_{\mathsf{F}}$$
$$\leq 2r\left\|\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\top}-\boldsymbol{Y}_1^{t,\ell}\boldsymbol{Y}_1^{t,\ell\top}\right\|_{\mathsf{F}}\cdot\left\|\mathcal{Z}^t\right\|,$$

where the fourth line is due to Lemma 22.

Combining $\xi_{2,4}^a$ and $\xi_{2,4}^b$ together gives that

$$\left\|\mathcal{M}_1\left(\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq 1}}-\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t,\ell},\{\boldsymbol{X}_j^{t,\ell}\}_{j\neq 1}}\right)(\mathcal{Z}^t)\right)\right\|_{\mathsf{F}}$$
$$\leq 4r\max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t-\boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}}\cdot\left\|\mathcal{Z}^t\right\|+2r\max_{i=1,2,3}\left\|\boldsymbol{Y}_i^t\boldsymbol{Y}_i^{t\top}-\boldsymbol{Y}_i^{t,\ell}\boldsymbol{Y}_i^{t,\ell\top}\right\|_{\mathsf{F}}\cdot\left\|\mathcal{Z}^t\right\|. \quad (81)$$

Applying the same argument as above shows that for $i = 2, 3$,

$$\begin{aligned} &\left\|\mathcal{M}_i\left(\left(\mathcal{P}^{(j\neq i)}_{\mathcal{G}^t,\{\boldsymbol{X}_j^t\}_{j\neq i}} - \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell},\{\boldsymbol{X}_j^{t,\ell}\}_{j\neq i}}\right)(\mathcal{Z}^t)\right)\right\|_{\mathsf{F}} \\ &\quad \leq 4r \max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}} \cdot \left\|\mathcal{Z}^t\right\| + 2r \max_{i=1,2,3}\left\|\boldsymbol{Y}_i^t\boldsymbol{Y}_i^{t\top} - \boldsymbol{Y}_i^{t,\ell}\boldsymbol{Y}_i^{t,\ell\top}\right\|_{\mathsf{F}} \cdot \left\|\mathcal{Z}^t\right\|. \end{aligned} \tag{82}$$

Plugging (81) and (82) into (80), we can obtain

$$\xi_{2,4} \leq 12r \max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}} \cdot \left\|\mathcal{Z}^t\right\| + 6r \max_{i=1,2,3}\left\|\boldsymbol{Y}_i^t\boldsymbol{Y}_i^{t\top} - \boldsymbol{Y}_i^{t,\ell}\boldsymbol{Y}_i^{t,\ell\top}\right\|_{\mathsf{F}} \cdot \left\|\mathcal{Z}^t\right\|. \tag{83}$$

Putting (75), (76), (79) and (83) together, one has

$$\begin{aligned} \xi_2 \leq &\left(8\left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}}\right)^3 + 12\left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}}\right)^2\right) \cdot \sqrt{n}\left\|\mathcal{Z}^t\right\| \\ &+ \left(24r^2\left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}\right\|_{\mathsf{F}}\right) + 12r\max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}}\right) \cdot \left\|\mathcal{Z}^t\right\| \\ &+ 6r\max_{i=1,2,3}\left\|\boldsymbol{Y}_i^t\boldsymbol{Y}_i^{t\top} - \boldsymbol{Y}_i^{t,\ell}\boldsymbol{Y}_i^{t,\ell\top}\right\|_{\mathsf{F}} \cdot \left\|\mathcal{Z}^t\right\|. \end{aligned}$$

By the induction hypothesis (17e), it can be seen that

$$\begin{aligned} 8\sqrt{n}\left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}}\right)^3 &\leq 8\sqrt{n}\left(\frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\right)^3 \leq \frac{1}{5\cdot 72}\sqrt{\frac{\mu r}{n}}, \\ 12\sqrt{n}\left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}}\right)^2 &\leq 12\sqrt{n}\left(\frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\right)^2 \leq \frac{1}{5\cdot 72}\sqrt{\frac{\mu r}{n}}, \\ 24r^2\left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top} - \boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}\right\|_{\mathsf{F}}\right) &\leq 24r^2\frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}} \leq \frac{1}{5\cdot 72}\sqrt{\frac{\mu r}{n}}, \\ 12r\max_{i=1,2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}} &\leq 12r\frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}} \leq \frac{1}{5\cdot 72}\sqrt{\frac{\mu r}{n}}. \end{aligned}$$

Moreover, Claim A.4 implies that

$$6r\max_{i=1,2,3}\left\|\boldsymbol{Y}_i^t\boldsymbol{Y}_i^{t\top} - \boldsymbol{Y}_i^{t,\ell}\boldsymbol{Y}_i^{t,\ell\top}\right\|_{\mathsf{F}} \leq 6r\cdot\frac{1}{2^{13}}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}} \leq \frac{r}{5\cdot 72}\sqrt{\frac{\mu r}{n}}.$$

Using these bounds, we can obtain

$$\begin{aligned} \xi_2 &\leq 5\cdot\frac{r}{5\cdot 72}\sqrt{\frac{\mu r}{n}}\left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t - \mathcal{T}\right)\right\| \\ &\leq 5\cdot\frac{r}{5\cdot 72}\sqrt{\frac{\mu r}{n}}\cdot 2r\cdot\left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right\|\left\|\mathcal{X}^t - \mathcal{T}\right\|_\infty \qquad (84) \\ &\leq 5\cdot\frac{r}{5\cdot 72}\sqrt{\frac{\mu r}{n}}\cdot C\left(\frac{\log^3 n}{p} + \sqrt{\frac{n\log^5 n}{p}}\right)\cdot 2r\cdot\frac{36}{2^{20}\kappa^2\mu^2r^4}\cdot\frac{1}{2^t}\cdot\left(\frac{\mu r}{n}\right)^{3/2}\cdot\sigma_{\max}(\mathcal{T}) \\ &\leq \frac{1}{2^8}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \qquad (85) \end{aligned}$$

provided that $p \geq \max\left\{\frac{C_1\kappa^2\mu^{1.5}r^{3.5}\log^3 n}{n^{1.5}}, \frac{C_2\kappa^4\mu^3r^7\log^5 n}{n^2}\right\}$.

Bounding ξ_3. For notational convenience, define

$$\mathcal{Z}_a = \left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t - \mathcal{X}^{t,\ell}\right) \quad \text{and} \quad \mathcal{Z}_b = \left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right).$$

The triangle inequality gives that

$$\begin{aligned}
\xi_3 &= \left\| \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t - \mathcal{T}\right) - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{I} - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right) \right\|_{\mathsf{F}} \\
&= \left\| \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^t - \mathcal{X}^{t,\ell}\right) + \left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right) \right\|_{\mathsf{F}} \\
&\le \left\| \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{Z}_a\right) \right\|_{\mathsf{F}} + \left\| \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{Z}_b\right) \right\|_{\mathsf{F}}.
\end{aligned}$$

By the definition of $\mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}$, the first term on the right hand side can be bounded as follows:

$$\begin{aligned}
\left\| \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{Z}_a\right) \right\|_{\mathsf{F}} &\le \left\| \prod_{i=1}^3 \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell}}(\mathcal{Z}_a) \right\|_{\mathsf{F}} + \sum_{i=1}^3 \left\| \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\ell,\perp}} \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell}, \left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}}(\mathcal{Z}_a) \right\|_{\mathsf{F}} \\
&\le \left\| \mathcal{Z}_a \underset{j\neq 1}{\times} \boldsymbol{X}_j^{t,\ell\mathsf{T}} \right\|_{\mathsf{F}} + \sum_{i=1}^3 \left\| \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell}, \left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}}(\mathcal{Z}_a) \right\|_{\mathsf{F}}. \qquad (86)
\end{aligned}$$

Moreover,

$$\begin{aligned}
\left\| \mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t,\ell}, \left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq 1}}(\mathcal{Z}_a) \right\|_{\mathsf{F}} &= \left\| \mathcal{M}_1\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^{t,\ell}, \left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq 1}}(\mathcal{Z}_a) \right) \right\|_{\mathsf{F}} \\
&= \left\| \mathcal{M}_1(\mathcal{Z}_a)\left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell}\right) \mathcal{M}_1^\dagger\left(\mathcal{G}^{t,\ell}\right) \mathcal{M}_1\left(\mathcal{G}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell}\right)^{\mathsf{T}} \right\|_{\mathsf{F}} \\
&\le \left\| \mathcal{M}_1(\mathcal{Z}_a)\left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell}\right) \right\|_{\mathsf{F}} = \left\| \mathcal{Z}_a \underset{j\neq 1}{\times} \boldsymbol{X}_j^{t,\ell\mathsf{T}} \right\|_{\mathsf{F}}. \qquad (87)
\end{aligned}$$

Similarly, one has

$$\left\| \mathcal{P}^{(j\neq i)}_{\mathcal{G}^{t,\ell}, \left\{\boldsymbol{X}_j^{t,\ell}\right\}_{j\neq i}}(\mathcal{Z}_a) \right\|_{\mathsf{F}} \le \left\| \mathcal{Z}_a \underset{j\neq i}{\times} \boldsymbol{X}_j^{t,\ell\mathsf{T}} \right\|_{\mathsf{F}}, \quad i = 2, 3. \qquad (88)$$

Plugging (87) and (88) into (86) yields that

$$\left\| \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{Z}_a\right) \right\|_{\mathsf{F}} \le 4 \max_{i=1,2,3} \left\| \mathcal{Z}_a \underset{j\neq i}{\times} \boldsymbol{X}_j^{t,\ell\mathsf{T}} \right\|_{\mathsf{F}}. \qquad (89)$$

Using the same argument, one can obtain

$$\left\| \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}\left(\mathcal{Z}_b\right) \right\|_{\mathsf{F}} \le 4 \max_{i=1,2,3} \left\| \mathcal{Z}_b \underset{j\neq i}{\times} \boldsymbol{X}_j^{t,\ell\mathsf{T}} \right\|_{\mathsf{F}}. \qquad (90)$$

Combining (89) and (90) together shows that

$$\xi_3 \le 4 \max_{i=1,2,3} \left\| \mathcal{Z}_a \underset{j\neq i}{\times} \boldsymbol{X}_j^{t,\ell\mathsf{T}} \right\|_{\mathsf{F}} + 4 \max_{i=1,2,3} \left\| \mathcal{Z}_b \underset{j\neq i}{\times} \boldsymbol{X}_j^{t,\ell\mathsf{T}} \right\|_{\mathsf{F}}.$$

To proceed, we first present the upper bounds of these terms in the following claim, whose proofs are deferred to Section C.5 and Section C.6, respectively.

Claim A.5 *Suppose* $p \geq \max\left\{\frac{C_1\kappa^2\mu^{1.5}r^{2.5}\log^3 n}{n^{1.5}}, \frac{C_2\kappa^4\mu^3 r^5\log^5 n}{n^2}\right\}$. *For any* $i = 1,2,3$, *the following inequalities hold with high probability*

$$\left\|\mathcal{Z}_a \underset{j\neq i}{\times} \boldsymbol{X}_j^{t,\ell\mathsf{T}}\right\|_{\mathsf{F}} \leq \frac{1}{2^{11}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \tag{91}$$

$$\left\|\mathcal{Z}_b \underset{j\neq i}{\times} \boldsymbol{X}_j^{t,\ell\mathsf{T}}\right\|_{\mathsf{F}} \leq \frac{1}{2^{11}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}). \tag{92}$$

It follows immediately from the claim that

$$\xi_3 \leq \frac{1}{2^8}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}). \tag{93}$$

Combining ξ_1, ξ_2 and ξ_3 together. Plugging (73), (84) and (93) into (67) yields that

$$\|\mathcal{E}^t - \mathcal{E}^{t,\ell}\|_{\mathsf{F}} \leq \frac{1}{2^6}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}). \tag{94}$$

A.2.4 Bounding $\left\|\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1} - \boldsymbol{X}_i^{t+1,\ell}\boldsymbol{T}_i^{t+1,\ell}\right\|_{\mathsf{F}}$

Recall that $\boldsymbol{X}_i^{t+1}\boldsymbol{\Sigma}_i^{t+1}\boldsymbol{X}_i^{t+1\mathsf{T}}$ and $\boldsymbol{X}_i^{t+1,\ell}\boldsymbol{\Sigma}_i^{t+1,\ell}\boldsymbol{X}_i^{t+1,\ell\mathsf{T}}$ are the top-r eigenvalue decomposition of $\left(\boldsymbol{T}_i + \boldsymbol{E}_i^t\right)\left(\boldsymbol{T}_i + \boldsymbol{E}_i^t\right)^\mathsf{T}$ and $\left(\boldsymbol{T}_i + \boldsymbol{E}_i^{t,\ell}\right)\left(\boldsymbol{T}_i + \boldsymbol{E}_i^{t,\ell}\right)^\mathsf{T}$, respectively. By the Weyl's inequality, the eigengap δ between the r-th and $r{+}1$-th eigenvalues of $\left(\boldsymbol{T}_i + \boldsymbol{E}_i^{t,\ell}\right)\left(\boldsymbol{T}_i + \boldsymbol{E}_i^{t,\ell}\right)^\mathsf{T}$ is bounded below as follows:

$$\begin{aligned}\delta &\geq \sigma_r(\boldsymbol{T}_i\boldsymbol{T}_i^\mathsf{T}) - 2\left\|\left(\boldsymbol{T}_i + \boldsymbol{E}_i^{t,\ell}\right)\left(\boldsymbol{T}_i + \boldsymbol{E}_i^{t,\ell}\right)^\mathsf{T} - \boldsymbol{T}_i\boldsymbol{T}_i^\mathsf{T}\right\| \\ &\geq \sigma_{\min}^2(\mathcal{T}) - 2\left\|\boldsymbol{T}_i\boldsymbol{E}_i^{t,\ell\mathsf{T}} + \boldsymbol{E}_i^{t,\ell}\boldsymbol{T}_i^\mathsf{T} + \boldsymbol{E}_i^{t,\ell}\boldsymbol{E}_i^{t,\ell\mathsf{T}}\right\| \geq \left(1 - \frac{1}{2^{20}}\right)\cdot\sigma_{\min}^2(\mathcal{T}),\end{aligned}$$

where the last inequality is due to (58). Define the perturbation matrix $\boldsymbol{W}_i$ by

$$\begin{aligned}\boldsymbol{W}_i &:= \left(\boldsymbol{T}_i + \boldsymbol{E}_i^t\right)\left(\boldsymbol{T}_i + \boldsymbol{E}_i^t\right)^\mathsf{T} - \left(\boldsymbol{T}_i + \boldsymbol{E}_i^{t,\ell}\right)\left(\boldsymbol{T}_i + \boldsymbol{E}_i^{t,\ell}\right)^\mathsf{T} \\ &= \left(\boldsymbol{T}_i + \boldsymbol{E}_i^t\right)\left(\boldsymbol{E}_i^t - \boldsymbol{E}_i^{t,\ell}\right)^\mathsf{T} + \left(\boldsymbol{E}_i^t - \boldsymbol{E}_i^{t,\ell}\right)\left(\boldsymbol{T}_i + \boldsymbol{E}_i^{t,\ell}\right)^\mathsf{T}.\end{aligned}$$

We have

$$\begin{aligned}\|\boldsymbol{W}_i\| \leq \|\boldsymbol{W}_i\|_{\mathsf{F}} &\leq \left\|\left(\boldsymbol{T}_i + \boldsymbol{E}_i^t\right)\left(\boldsymbol{E}_i^t - \boldsymbol{E}_i^{t,\ell}\right)^\mathsf{T}\right\|_{\mathsf{F}} + \left\|\left(\boldsymbol{E}_i^t - \boldsymbol{E}_i^{t,\ell}\right)\left(\boldsymbol{T}_i + \boldsymbol{E}_i^{t,\ell}\right)^\mathsf{T}\right\|_{\mathsf{F}} \\ &\leq \left(\left\|\boldsymbol{T}_i + \boldsymbol{E}_i^t\right\| + \left\|\boldsymbol{T}_i + \boldsymbol{E}_i^{t,\ell}\right\|\right)\left\|\mathcal{E}^t - \mathcal{E}^{t,\ell}\right\|_{\mathsf{F}} \leq \frac{5}{2}\sigma_{\max}(\mathcal{T})\left\|\mathcal{E}^t - \mathcal{E}^{t,\ell}\right\|_{\mathsf{F}} \\ &\leq \frac{5}{2}\sigma_{\max}(\mathcal{T})\cdot\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\cdot\sigma_{\max}(\mathcal{T}) \leq \frac{1}{2^{18}}\sigma_{\min}^2(\mathcal{T}).\end{aligned} \tag{95}$$

Applying Lemma 34 and Lemma 35 yields that

$$\begin{aligned}\left\|\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1}-\boldsymbol{X}_i^{t+1,\ell}\boldsymbol{T}_i^{t+1,\ell}\right\|_{\mathsf{F}} &\le \frac{\sqrt{2}\left\|\boldsymbol{W}_i\right\|_{\mathsf{F}}}{\delta-\left\|\boldsymbol{W}_i\right\|} \le 2\frac{\left\|\boldsymbol{W}_i\right\|_{\mathsf{F}}}{\sigma_{\min}^2(\mathcal{T})}\\ &\overset{(a)}{\le} \frac{2}{\sigma_{\min}^2(\mathcal{T})}\cdot\frac{5}{2}\sigma_{\max}(\mathcal{T})\cdot\left\|\mathcal{E}^t-\mathcal{E}^{t,\ell}\right\|_{\mathsf{F}}\\ &\overset{(b)}{\le} \frac{2}{\sigma_{\min}^2(\mathcal{T})}\cdot\frac{5}{2}\sigma_{\max}(\mathcal{T})\cdot\frac{1}{2^6}\cdot\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T})\\ &\le \frac{1}{2^3}\frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}},\end{aligned} \tag{96}$$

where (a) follows from (95) and (b) is due to (94).

A.2.5 Bounding $\left\|\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1}-\boldsymbol{U}_i\right\|_{2,\infty}$

A direct computation yields that

$$\begin{aligned}\left\|\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1}-\boldsymbol{U}_i\right\|_{2,\infty} &\le \left\|\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1}-\boldsymbol{X}_i^{t+1,\ell}\boldsymbol{R}_i^{t+1,\ell}\right\|_{2,\infty}+\left\|\boldsymbol{X}_i^{t+1,\ell}\boldsymbol{R}_i^{t+1,\ell}-\boldsymbol{U}_i\right\|_{2,\infty}\\ &\le \left\|\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1}-\boldsymbol{X}_i^{t+1,\ell}\boldsymbol{R}_i^{t+1,\ell}\right\|_{\mathsf{F}}+\left\|\boldsymbol{X}_i^{t+1,\ell}\boldsymbol{R}_i^{t+1,\ell}-\boldsymbol{U}_i\right\|_{2,\infty}.\end{aligned} \tag{97}$$

We will use Lemma 32 with $\boldsymbol{X}_1=\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1}$ and $\boldsymbol{X}_2=\boldsymbol{X}_i^{t+1,\ell}\boldsymbol{T}_i^{t+1,\ell}$ to bound the first term in (97). Following the same argument as in the proof of Lemma 18, one can obtain

$$\left\|\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1}-\boldsymbol{U}_i\right\|\left\|\boldsymbol{U}_i\right\| \le \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}} \le \frac{1}{2}.$$

Furthermore, the inequality (96) shows that

$$\left\|\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1}-\boldsymbol{X}_i^{t+1,\ell}\boldsymbol{T}_i^{t+1,\ell}\right\|\left\|\boldsymbol{U}_i\right\| \le \frac{1}{2^3}\frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^{t+1}} \le \frac{1}{4}.$$

Thus, by Lemma 32, one has

$$\begin{aligned}\left\|\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1}-\boldsymbol{X}_i^{t+1,\ell}\boldsymbol{R}_i^{t+1,\ell}\right\|_{\mathsf{F}} &\le 5\left\|\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1}-\boldsymbol{X}_i^{t+1,\ell}\boldsymbol{T}_i^{t+1,\ell}\right\|_{\mathsf{F}}\\ &\le \frac{5}{2^3}\frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^{t+1}}.\end{aligned} \tag{98}$$

Plugging (66) and (98) into (97) shows that

$$\left\|\boldsymbol{X}_i^{t+1}\boldsymbol{R}_i^{t+1}-\boldsymbol{U}_i\right\|_{2,\infty} \le \frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}.$$

Appendix B. Proofs of Key Lemmas

B.1 Proof of Lemma 5

The $\ell_{2,\infty}$ norm of $\boldsymbol{T}_1$ can be bounded as follows

$$\left\|\boldsymbol{T}_1\right\|_{2,\infty}=\left\|\boldsymbol{U}_1\mathcal{M}_1(\mathcal{S})(\boldsymbol{U}_3\otimes\boldsymbol{U}_2)^{\mathsf{T}}\right\|_{2,\infty} \le \left\|\boldsymbol{U}_1\right\|_{2,\infty}\left\|\mathcal{M}_1(\mathcal{S})\right\| \le \sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}).$$

For $\left\|\boldsymbol{T}_1^\top\right\|_{2,\infty}$, a direct calculation yields that

$$\left\|\boldsymbol{T}_1^\top\right\|_{2,\infty} = \left\|(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2)\mathcal{M}_1^\top(\mathcal{S})\boldsymbol{U}_1^\top\right\|_{2,\infty} \le \|\boldsymbol{U}_2\|_{2,\infty} \cdot \|\boldsymbol{U}_3\|_{2,\infty} \cdot \sigma_{\max}(\mathcal{T}) \le \frac{\mu r}{n}\sigma_{\max}(\mathcal{T}),$$

where the second line following from the fact $\|\boldsymbol{U}_3 \otimes \boldsymbol{U}_2\|_{2,\infty} \le \|\boldsymbol{U}_2\|_{2,\infty} \cdot \|\boldsymbol{U}_3\|_{2,\infty}$. Using the same argument, one can obtain

$$\|\boldsymbol{T}_i\|_{2,\infty} \le \sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}) \quad \text{and} \quad \left\|\boldsymbol{T}_i^\top\right\|_{2,\infty} \le \frac{\mu r}{n}\sigma_{\max}(\mathcal{T}), \quad i = 2, 3.$$

Lastly, by the definition of tensor infinity norm, one has

$$\begin{aligned}\|\mathcal{T}\|_\infty = \|\mathcal{M}_1(\mathcal{T})\|_\infty &= \left\|\boldsymbol{U}_1\mathcal{M}_1(\mathcal{S})\left(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2\right)^\top\right\|_\infty \\ &\le \prod_{j=1}^{3}\|\boldsymbol{U}_j\|_{2,\infty} \cdot \|\mathcal{M}_1(\mathcal{S})\| \le \left(\frac{\mu r}{n}\right)^{3/2}\sigma_{\max}(\mathcal{T}).\end{aligned}$$

B.2 Proof of Lemma 10

Notice that $\boldsymbol{X}_i$ is the top-r singular vectors of

$$\left(\boldsymbol{T}_i + \boldsymbol{E}_i\right)\left(\boldsymbol{T}_i + \boldsymbol{E}_i\right)^\top = \boldsymbol{T}_i\boldsymbol{T}_i^\top + \underbrace{\boldsymbol{T}_i\boldsymbol{E}_i^\top + \boldsymbol{E}_i\boldsymbol{T}_i^\top + \boldsymbol{E}_i\boldsymbol{E}_i^\top}_{=:\boldsymbol{\Delta}_i},$$

where $\boldsymbol{T}_i = \mathcal{M}_i\left(\mathcal{T}\right)$ and $\boldsymbol{E}_i = \mathcal{M}_i\left(\mathcal{E}\right)$. Let $\boldsymbol{T}_i\boldsymbol{T}_i^\top = \boldsymbol{U}_i\boldsymbol{\Lambda}_i\boldsymbol{U}_i^\top$ be the eigenvalue decomposition. We will apply Lemma 26 to bound $\|\boldsymbol{X}_i\boldsymbol{R}_i - \boldsymbol{U}_i\|$ where $\boldsymbol{R}_i = \arg\min_{\boldsymbol{R}^\top\boldsymbol{R}=\boldsymbol{I}}\|\boldsymbol{X}_i\boldsymbol{R} - \boldsymbol{U}_i\|_\mathsf{F}$. Invoking the triangle inequality shows that

$$\|\boldsymbol{\Delta}_i\| \le \left(2\|\boldsymbol{T}_i\| + \|\boldsymbol{E}_i\|\right)\|\boldsymbol{E}_i\| \le \frac{5}{2}\sigma_{\max}\left(\mathcal{T}\right)\|\boldsymbol{E}_i\| \le \frac{5}{2}\sigma_{\max}\left(\mathcal{T}\right) \cdot \frac{1}{10\kappa^2}\sigma_{\max}\left(\mathcal{T}\right) \le \frac{1}{4}\sigma_{\min}^2\left(\mathcal{T}\right),$$

where the second and third lines follow from the assumption $\max_{i=1,2,3}\|\boldsymbol{E}_i\| \le \frac{\sigma_{\max}(\mathcal{T})}{10\kappa^2} \le \frac{\sigma_{\max}(\mathcal{T})}{2}$. Thus, the requirement in Lemma 26 is valid, and we have

$$\|\boldsymbol{X}_i\boldsymbol{R}_i - \boldsymbol{U}_i\| \le \frac{3}{\sigma_{\min}^2(\mathcal{T})}\|\boldsymbol{\Delta}_i\| \le \frac{15\sigma_{\max}(\mathcal{T})}{2\sigma_{\min}^2(\mathcal{T})}\|\boldsymbol{E}_i\|.$$

Furthermore, $\|\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{U}_i\boldsymbol{U}_i^\top\|_{2,\infty}$ can be bounded as follows:

$$\begin{aligned}\left\|\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{U}_i\boldsymbol{U}_i^\top\right\|_{2,\infty} &\le \left\|(\boldsymbol{X}_i\boldsymbol{R}_i - \boldsymbol{U}_i)(\boldsymbol{X}_i\boldsymbol{R}_i)^\top\right\|_{2,\infty} + \left\|\boldsymbol{U}_i(\boldsymbol{X}_i\boldsymbol{R}_i - \boldsymbol{U}_i)^\top\right\|_{2,\infty} \\ &\le \|\boldsymbol{X}_i\boldsymbol{R}_i - \boldsymbol{U}_i\|_{2,\infty} + \|\boldsymbol{U}_i\|_{2,\infty} \cdot \|\boldsymbol{X}_i\boldsymbol{R}_i - \boldsymbol{U}_i\| \\ &\le \|\boldsymbol{X}_i\boldsymbol{R}_i - \boldsymbol{U}_i\|_{2,\infty} + \frac{15\sigma_{\max}(\mathcal{T})}{2\sigma_{\min}^2(\mathcal{T})}\|\boldsymbol{E}_i\| \cdot \|\boldsymbol{U}_i\|_{2,\infty} \\ &\le \max_{i=1,2,3}\left(\|\boldsymbol{X}_i\boldsymbol{R}_i - \boldsymbol{U}_i\|_{2,\infty} + \frac{15\sigma_{\max}(\mathcal{T})}{2\sigma_{\min}^2(\mathcal{T})}\|\boldsymbol{E}_i\|\,\|\boldsymbol{U}_i\|_{2,\infty}\right) =: B.\end{aligned} \tag{99}$$

Next, we turn to control $\|\mathcal{X}-\mathcal{T}\|_\infty$,

$$\begin{aligned}
\|\mathcal{X}-\mathcal{T}\|_\infty &= \left\|(\mathcal{T}+\mathcal{E}) \bigtimes_{i=1}^{3} \boldsymbol{X}_i\boldsymbol{X}_i^\top - \mathcal{T}\right\|_\infty \\
&= \left\|\mathcal{T} \bigtimes_{i=1}^{3} \boldsymbol{X}_i\boldsymbol{X}_i^\top - \mathcal{T} \bigtimes_{i=1}^{3} \boldsymbol{U}_i\boldsymbol{U}_i^\top + \mathcal{E} \bigtimes_{i=1}^{3} \boldsymbol{X}_i\boldsymbol{X}_i^\top\right\|_\infty \\
&\le \left\|\mathcal{T} \bigtimes_{i=1}^{3} \boldsymbol{X}_i\boldsymbol{X}_i^\top - \mathcal{T} \bigtimes_{i=1}^{3} \boldsymbol{U}_i\boldsymbol{U}_i^\top\right\|_\infty + \left\|\mathcal{E} \bigtimes_{i=1}^{3} \boldsymbol{X}_i\boldsymbol{X}_i^\top\right\|_\infty \\
&\le \underbrace{\left\|\mathcal{T} \bigtimes_{i=1}^{3} \left(\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{U}_i\boldsymbol{U}_i^\top\right)\right\|_\infty}_{=:\omega_1} + \underbrace{\sum_{i=1}^{3}\left\|\mathcal{T} \times_i \boldsymbol{U}_i\boldsymbol{U}_i^\top \bigtimes_{j\neq i} \left(\boldsymbol{X}_j\boldsymbol{X}_j^\top - \boldsymbol{U}_j\boldsymbol{U}_j^\top\right)\right\|_\infty}_{=:\omega_2} \\
&\quad + \underbrace{\sum_{i=1}^{3}\left\|\mathcal{T} \times_i \left(\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{U}_i\boldsymbol{U}_i^\top\right) \bigtimes_{j\neq i} \boldsymbol{U}_j\boldsymbol{U}_j^\top\right\|_\infty}_{=:\omega_3} + \underbrace{\left\|\mathcal{E} \bigtimes_{i=1}^{3} \boldsymbol{X}_i\boldsymbol{X}_i^\top\right\|_\infty}_{=:\omega_4}.
\end{aligned}$$

Bounding ω_1. It is easy to see that

$$\begin{aligned}
\omega_1 &= \left\|\mathcal{M}_1\left(\mathcal{T} \bigtimes_{i=1}^{3} \left(\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{U}_i\boldsymbol{U}_i^\top\right)\right)\right\|_\infty \\
&\le \left\|\boldsymbol{X}_1\boldsymbol{X}_1^\top - \boldsymbol{U}_1\boldsymbol{U}_1^\top\right\|_{2,\infty} \cdot \|\boldsymbol{T}_1\| \cdot \left\|\left(\boldsymbol{X}_3\boldsymbol{X}_3^\top - \boldsymbol{U}_3\boldsymbol{U}_3^\top\right) \otimes \left(\boldsymbol{X}_2\boldsymbol{X}_2^\top - \boldsymbol{U}_2\boldsymbol{U}_2^\top\right)\right\|_{2,\infty} \\
&\le \sigma_{\max}(\mathcal{T}) \cdot \prod_{i=1}^{3}\left\|\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{U}_i\boldsymbol{U}_i^\top\right\|_{2,\infty} \le B^3 \cdot \sigma_{\max}(\mathcal{T}),
\end{aligned}$$

where the last inequality is due to (99).

Bounding ω_2. A directly calculation gives

$$\begin{aligned}
\omega_2 &= \sum_{i=1}^{3}\left\|\mathcal{M}_i\left(\mathcal{T} \times_i \boldsymbol{U}_i\boldsymbol{U}_i^\top \bigtimes_{j\neq i} \left(\boldsymbol{X}_j\boldsymbol{X}_j^\top - \boldsymbol{U}_j\boldsymbol{U}_j^\top\right)\right)\right\|_\infty \\
&\le \sum_{i=1}^{3} \|\boldsymbol{U}_i\|_{2,\infty} \|\boldsymbol{T}_i\| \prod_{j\neq i}\left\|\boldsymbol{X}_j\boldsymbol{X}_j^\top - \boldsymbol{U}_j\boldsymbol{U}_j^\top\right\|_{2,\infty} \le 3 \max_{i=1,2,3}\|\boldsymbol{U}_i\|_{2,\infty} \cdot \sigma_{\max}(\mathcal{T}) \cdot B^2.
\end{aligned}$$

Bounding ω_3. A straightforward computation shows that

$$\omega_3 = \sum_{i=1}^{3}\left\|\mathcal{T} \times_i \left(\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{U}_i\boldsymbol{U}_i^\top\right) \bigtimes_{j\neq i} \boldsymbol{U}_j\boldsymbol{U}_j^\top\right\|_\infty \le 3\left(\max_{i=1,2,3}\|\boldsymbol{U}_i\|_{2,\infty}\right)^2 \cdot \sigma_{\max}(\mathcal{T}) \cdot B.$$

Bounding ω_4. By the definition of the infinity norm, one has

$$\begin{aligned}
\omega_4 &= \sup_{\boldsymbol{e}_{i_1},\boldsymbol{e}_{i_2},\boldsymbol{e}_{i_3}} \left|\mathcal{E} \times_1 \boldsymbol{e}_{i_1}^\top \boldsymbol{X}_1\boldsymbol{X}_1^\top \times_2 \boldsymbol{e}_{i_2}^\top \boldsymbol{X}_2\boldsymbol{X}_2^\top \times_3 \boldsymbol{e}_{i_3}^\top \boldsymbol{X}_3\boldsymbol{X}_3^\top\right| \\
&\le \|\mathcal{E}\| \cdot \left(\max_{i=1,2,3}\|\boldsymbol{X}_i\|_{2,\infty}\right)^3 \overset{(a)}{\le} \|\mathcal{M}_1(\mathcal{E})\| \cdot \left(\max_{i=1,2,3}\|\boldsymbol{X}_i\|_{2,\infty}\right)^3.
\end{aligned}$$

where step (a) follows from the fact $\|\mathcal{X}\| \le \|\mathcal{M}_i(\mathcal{X})\|$ for any tensor $\mathcal{X}$.

Combining these four upper bounds together, we complete the proof.

B.3 Proof of Lemma 11

By the definition of operator $\mathcal{P}_T$, one can obtain

$$\|\mathcal{P}_T(\boldsymbol{e}_{i_1} \circ \boldsymbol{e}_{i_2} \circ \boldsymbol{e}_{i_3})\|_\mathsf{F}^2 = \left\|(\boldsymbol{e}_{i_1} \circ \boldsymbol{e}_{i_2} \circ \boldsymbol{e}_{i_3}) \mathop{\times}_{i=1}^{3} \boldsymbol{U}_i\boldsymbol{U}_i^\mathsf{T}\right\|_\mathsf{F}^2 + \sum_{i=1}^{3} \left\|\mathcal{S} \times_i \boldsymbol{W}_i \mathop{\times}_{j\neq i} \boldsymbol{U}_j\right\|_\mathsf{F}^2 =: \psi_1 + \psi_2.$$

Bounding ψ_1. Recall the definition of $\boldsymbol{e}_{i_1} \circ \boldsymbol{e}_{i_2} \circ \boldsymbol{e}_{i_3}$ in (4). One has

$$\begin{aligned}
\psi_1 &= \left\|\boldsymbol{U}_1\boldsymbol{U}_1^\mathsf{T}\mathcal{M}_1(\boldsymbol{e}_{i_1} \circ \boldsymbol{e}_{i_2} \circ \boldsymbol{e}_{i_3})\left(\boldsymbol{U}_3\boldsymbol{U}_3^\mathsf{T} \otimes \boldsymbol{U}_2\boldsymbol{U}_2^\mathsf{T}\right)^\mathsf{T}\right\|_\mathsf{F}^2 \\
&= \left\|\boldsymbol{U}_1\boldsymbol{U}_1^\mathsf{T}\boldsymbol{e}_{i_1}(\boldsymbol{e}_{i_3} \otimes \boldsymbol{e}_{i_2})^\mathsf{T}\left(\boldsymbol{U}_3\boldsymbol{U}_3^\mathsf{T} \otimes \boldsymbol{U}_2\boldsymbol{U}_2^\mathsf{T}\right)^\mathsf{T}\right\|_\mathsf{F}^2 \\
&= \left\|\boldsymbol{U}_1\boldsymbol{U}_1^\mathsf{T}\boldsymbol{e}_{i_1}\left(\boldsymbol{U}_3\boldsymbol{U}_3^\mathsf{T}\boldsymbol{e}_{i_3} \otimes \boldsymbol{U}_2\boldsymbol{U}_2^\mathsf{T}\boldsymbol{e}_{i_2}\right)^\mathsf{T}\right\|_\mathsf{F}^2 \\
&\le \left\|\boldsymbol{U}_1^\mathsf{T}\boldsymbol{e}_{i_1}\right\|_2^2 \cdot \left\|\boldsymbol{U}_2^\mathsf{T}\boldsymbol{e}_{i_2}\right\|_2^2 \cdot \left\|\boldsymbol{U}_3^\mathsf{T}\boldsymbol{e}_{i_3}\right\|_2^2 \le \left(\frac{\mu r}{n}\right)^3.
\end{aligned}$$

Bounding ψ_2. We only provide details for bounding $\|\mathcal{S} \times_1 \boldsymbol{W}_1 \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{U}_3\|_\mathsf{F}^2$,

$$\begin{aligned}
&\|\mathcal{S} \times_1 \boldsymbol{W}_1 \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{U}_3\|_\mathsf{F}^2 \\
&\quad= \left\|\boldsymbol{W}_1\mathcal{M}_1(\mathcal{S})(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2)^\mathsf{T}\right\|_\mathsf{F}^2 \\
&\quad= \left\|\left(\boldsymbol{I} - \boldsymbol{U}_1\boldsymbol{U}_1^\mathsf{T}\right)\mathcal{M}_1(\boldsymbol{e}_{i_1} \circ \boldsymbol{e}_{i_2} \circ \boldsymbol{e}_{i_3})(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2)\mathcal{M}_1(\mathcal{S})^\dagger\mathcal{M}_1(\mathcal{S})(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2)^\mathsf{T}\right\|_\mathsf{F}^2 \\
&\quad\le \left\|(\boldsymbol{e}_{i_3} \otimes \boldsymbol{e}_{i_2})^\mathsf{T}(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2)\right\|_2^2 \cdot \left\|\mathcal{M}_1(\mathcal{S})^\dagger\mathcal{M}_1(\mathcal{S})\right\|^2 \le \left\|\boldsymbol{U}_2^\mathsf{T}\boldsymbol{e}_{i_2}\right\|_2^2 \cdot \left\|\boldsymbol{U}_3^\mathsf{T}\boldsymbol{e}_{i_3}\right\|_2^2 \le \left(\frac{\mu r}{n}\right)^2.
\end{aligned}$$

Similarly, one can obtain

$$\|\mathcal{S} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{W}_2 \times_3 \boldsymbol{U}_3\|_\mathsf{F}^2 \le \left(\frac{\mu r}{n}\right)^2 \text{ and } \|\mathcal{S} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{W}_3\|_\mathsf{F}^2 \le \left(\frac{\mu r}{n}\right)^2.$$

Therefore, $\psi_2 \le 3\left(\frac{\mu r}{n}\right)^2$.

Combining the above bounds together shows that

$$\|\mathcal{P}_T(\boldsymbol{e}_{i_1} \circ \boldsymbol{e}_{i_2} \circ \boldsymbol{e}_{i_3})\|_\mathsf{F}^2 \le 4\left(\frac{\mu r}{n}\right)^2.$$

B.4 Proof of Lemma 12

Let $\mathcal{E}_{i_1,i_2,i_3}$ be the tensor with the (i_1, i_2, i_3)-th entry being 1 and all the other entries being 0. The operator norm $\left\|\mathcal{P}_T\left(p^{-1}\mathcal{P}_\Omega - \mathcal{I}\right)\mathcal{P}_T\right\|$ can be expressed as follows:

$$\left\|\mathcal{P}_T\left(p^{-1}\mathcal{P}_\Omega - \mathcal{I}\right)\mathcal{P}_T\right\|$$

$$
\begin{aligned}
&= \sup_{\|\mathcal{Z}\|_{\mathsf{F}}=1, \mathcal{Z}\in\mathbb{R}^{n\times n\times n}} \left\| \mathcal{P}_T\left(p^{-1}\mathcal{P}_\Omega - \mathcal{I}\right)\mathcal{P}_T\left(\mathcal{Z}\right)\right\|_{\mathsf{F}} \\
&= \sup_{\|\mathcal{Z}\|_{\mathsf{F}}=1, \mathcal{Z}\in\mathbb{R}^{n\times n\times n}} \left\| \sum_{i_1,i_2,i_3} \left(p^{-1}\delta_{i_1,i_2,i_3} - 1\right) \left\langle \mathcal{P}_T\left(\mathcal{Z}\right), \mathcal{E}_{i_1,i_2,i_3}\right\rangle \mathcal{P}_T\left(\mathcal{E}_{i_1,i_2,i_3}\right) \right\|_{\mathsf{F}} \\
&= \sup_{\|\mathcal{Z}\|_{\mathsf{F}}=1, \mathcal{Z}\in\mathbb{R}^{n\times n\times n}} \left\| \sum_{i_1,i_2,i_3} \left(p^{-1}\delta_{i_1,i_2,i_3} - 1\right) \left\langle \mathcal{Z}, \mathcal{P}_T\left(\mathcal{E}_{i_1,i_2,i_3}\right)\right\rangle \mathrm{vec}\left(\mathcal{P}_T\left(\mathcal{E}_{i_1,i_2,i_3}\right)\right) \right\| \\
&= \sup_{\|\mathcal{Z}\|_{\mathsf{F}}=1, \mathcal{Z}\in\mathbb{R}^{n\times n\times n}} \left\| \sum_{i_1,i_2,i_3} \left(p^{-1}\delta_{i_1,i_2,i_3} - 1\right) \left\langle \mathrm{vec}(\mathcal{Z}), \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))\right\rangle \mathrm{vec}\left(\mathcal{P}_T\left(\mathcal{E}_{i_1,i_2,i_3}\right)\right) \right\| \\
&= \sup_{\|\mathcal{Z}\|_{\mathsf{F}}=1, \mathcal{Z}\in\mathbb{R}^{n\times n\times n}} \left\| \sum_{i_1,i_2,i_3} \left(p^{-1}\delta_{i_1,i_2,i_3} - 1\right) \mathrm{vec}\left(\mathcal{P}_T\left(\mathcal{E}_{i_1,i_2,i_3}\right)\right) \mathrm{vec}\left(\mathcal{P}_T\left(\mathcal{E}_{i_1,i_2,i_3}\right)\right)^{\mathsf{T}} \mathrm{vec}\left(\mathcal{Z}\right) \right\| \\
&= \left\| \sum_{i_1,i_2,i_3} \left(p^{-1}\delta_{i_1,i_2,i_3} - 1\right) \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}) \,\mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))^{\mathsf{T}} \right\|,
\end{aligned}
$$

where $\mathrm{vec}(\cdot)$ is used to denote the vectorization of a tensor. Define

$$
\boldsymbol{S}_{i_1,i_2,i_3} := \left(p^{-1}\delta_{i_1,i_2,i_3} - 1\right) \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))\, \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))^{\mathsf{T}} \in \mathbb{R}^{n^3\times n^3},
$$

which are independent symmetric matrices with mean zero. Thus we can apply the matrix Bernstein inequality to bound the above term. Lemma 11 gives that

$$
\|\boldsymbol{S}_{i_1,i_2,i_3}\| \le p^{-1} \left\|\mathcal{P}_T\left(\mathcal{E}_{i_1,i_2,i_3}\right)\right\|_{\mathsf{F}}^2 \le \frac{4}{p}\left(\frac{\mu r}{n}\right)^2 .
$$

Moreover, the operator norm of $\sum_{i_1,i_2,i_3} \mathbb{E}\left[\boldsymbol{S}_{i_1,i_2,i_3}\boldsymbol{S}_{i_1,i_2,i_3}^{\mathsf{T}}\right]$ can be bounded as follows:

$$
\begin{aligned}
&\left\| \sum_{i_1,i_2,i_3} \mathbb{E}\left[\boldsymbol{S}_{i_1,i_2,i_3}\boldsymbol{S}_{i_1,i_2,i_3}^{\mathsf{T}}\right] \right\| \\
&\le p^{-1} \left\| \sum_{i_1,i_2,i_3} \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))\, \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))^{\mathsf{T}}\, \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))\, \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))^{\mathsf{T}} \right\| \\
&\le p^{-1} \max_{i_1,i_2,i_3} \|\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3})\|_{\mathsf{F}}^2 \left\| \sum_{i_1,i_2,i_3} \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))\, \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))^{\mathsf{T}} \right\| \le \frac{4}{p}\left(\frac{\mu r}{n}\right)^2,
\end{aligned}
$$

where the least inequality follows from Lemma 11 and the fact

$$
\left\| \sum_{i_1,i_2,i_3} \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))\, \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))^{\mathsf{T}} \right\|
$$

$$
\begin{aligned}
&= \sup_{\|\mathcal{Z}\|_{\mathsf{F}}=1, \mathcal{Z}\in\mathbb{R}^{n\times n\times n}} \left\| \sum_{i_1,i_2,i_3} \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))\, \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))^{\mathsf{T}}\, \mathrm{vec}(\mathcal{Z}) \right\|_2 \\
&= \sup_{\|\mathcal{Z}\|_{\mathsf{F}}=1, \mathcal{Z}\in\mathbb{R}^{n\times n\times n}} \left\| \sum_{i_1,i_2,i_3} \langle \mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}), \mathcal{Z}\rangle \mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}) \right\|_{\mathsf{F}} \\
&= \sup_{\|\mathcal{Z}\|_{\mathsf{F}}=1, \mathcal{Z}\in\mathbb{R}^{n\times n\times n}} \left\| \mathcal{P}_T \left(\sum_{i_1,i_2,i_3} \langle \mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}), \mathcal{Z}\rangle \mathcal{E}_{i_1,i_2,i_3} \right) \right\|_{\mathsf{F}} \\
&\le \sup_{\|\mathcal{Z}\|_{\mathsf{F}}=1, \mathcal{Z}\in\mathbb{R}^{n\times n\times n}} \left\| \sum_{i_1,i_2,i_3} \langle \mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}), \mathcal{Z}\rangle \mathcal{E}_{i_1,i_2,i_3} \right\|_{\mathsf{F}} \\
&= \sup_{\|\mathcal{Z}\|_{\mathsf{F}}=1, \mathcal{Z}\in\mathbb{R}^{n\times n\times n}} \sqrt{\sum_{i_1,i_2,i_3} \langle \mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}), \mathcal{Z}\rangle^2} \\
&= \sup_{\|\mathcal{Z}\|_{\mathsf{F}}=1, \mathcal{Z}\in\mathbb{R}^{n\times n\times n}} \sqrt{\sum_{i_1,i_2,i_3} \langle \mathcal{E}_{i_1,i_2,i_3}, \mathcal{P}_T(\mathcal{Z})\rangle^2} \le 1.
\end{aligned}
$$

The application of Bernstein inequality yields that with high probability,

$$
\begin{aligned}
&\left\| \sum_{i_1,i_2,i_3} \left(p^{-1}\delta_{i_1,i_2,i_3} - 1\right) \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3})\, \mathrm{vec}(\mathcal{P}_T(\mathcal{E}_{i_1,i_2,i_3}))^{\mathsf{T}} \right\| \\
&\quad \le C\left(\frac{\log n}{p}\left(\frac{\mu r}{n}\right)^2 + \sqrt{\frac{\log n}{p}\left(\frac{\mu r}{n}\right)^2} \right) \le \varepsilon,
\end{aligned}
$$

provided $p \ge \frac{C_2\mu^2 r^2 \log n}{\varepsilon \cdot n^2}$, where ε is a small absolute constant.

B.5 Proof of Lemma 14

Applying Lemma 33 gives that for $i = 1,2,3$,

$$
\left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\mathsf{T}} - \boldsymbol{U}_i \boldsymbol{U}_i^{\mathsf{T}} \right\| \le \frac{\left\| \mathcal{M}_i\left(\mathcal{X}^t\right) - \mathcal{M}_i\left(\mathcal{T}\right)\right\|_{\mathsf{F}}}{\sigma_{\min}\left(\mathcal{M}_i\left(\mathcal{T}\right)\right)} \le \frac{\left\|\mathcal{X}^t - \mathcal{T}\right\|_{\mathsf{F}}}{\sigma_{\min}\left(\mathcal{T}\right)}.
$$

To prove the second inequality of Lemma 14, first note that for any tensor $\mathcal{Z} \in \mathbb{R}^{n\times n\times n}$,

$$
\left(\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_T\right)(\mathcal{Z}) = \prod_{i=1}^3 \mathcal{P}_{\boldsymbol{X}_i^t}^{(i)}(\mathcal{Z}) - \prod_{i=1}^3 \mathcal{P}_{\boldsymbol{U}_i}^{(i)}(\mathcal{Z}) + \sum_{i=1}^3 \left(\mathcal{P}^{(j\neq i)}_{\mathcal{G}^t, \left\{\boldsymbol{X}_j^t\right\}_{j\neq i}} \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\perp}} - \mathcal{P}^{(j\neq i)}_{\mathcal{S}, \{\boldsymbol{U}_j\}_{j\neq i}} \mathcal{P}^{(i)}_{\boldsymbol{U}_i{}^{\perp}} \right)(\mathcal{Z}).
$$

Taking the Frobenius norm on both sides and applying the triangle inequality give

$$
\left\|\left(\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_T\right)(\mathcal{Z})\right\|_{\mathsf{F}} \le \sum_{i=1}^3 \underbrace{\left\| \left(\mathcal{P}^{(j\neq i)}_{\mathcal{G}^t, \left\{\boldsymbol{X}_j^t\right\}_{j\neq i}} \mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\perp}} - \mathcal{P}^{(j\neq i)}_{\mathcal{S}, \{\boldsymbol{U}_j\}_{j\neq i}} \mathcal{P}^{(i)}_{\boldsymbol{U}_i{}^{\perp}} \right)(\mathcal{Z}) \right\|_{\mathsf{F}}}_{=:\varsigma_i} + \underbrace{\left\| \prod_{i=1}^3 \mathcal{P}_{\boldsymbol{X}_i^t}^{(i)}(\mathcal{Z}) - \prod_{i=1}^3 \mathcal{P}_{\boldsymbol{U}_i}^{(i)}(\mathcal{Z}) \right\|_{\mathsf{F}}}_{=:\varsigma_4}.
$$

Bounding ς_1. It is easy to see that

$$\begin{aligned}
\varsigma_1 &= \left\| \left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t, \{\boldsymbol{X}_j^t\}_{j\neq 1}} \mathcal{P}^{(1)}_{\boldsymbol{X}_1^{t,\perp}} - \mathcal{P}^{(j\neq 1)}_{\mathcal{S}, \{\boldsymbol{U}_j\}_{j\neq 1}} \mathcal{P}^{(1)}_{\boldsymbol{U}_1^{\perp}} \right)(\mathcal{Z}) \right\|_{\mathsf{F}} \\
&\leq \left\| \left(\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t, \{\boldsymbol{X}_j^t\}_{j\neq 1}} - \mathcal{P}^{(j\neq 1)}_{\mathcal{S}, \{\boldsymbol{U}_j\}_{j\neq 1}} \right) \mathcal{P}^{(1)}_{\boldsymbol{X}_1^{t,\perp}} \right)(\mathcal{Z}) \right\|_{\mathsf{F}} + \left\| \left(\mathcal{P}^{(j\neq 1)}_{\mathcal{S}, \{\boldsymbol{U}_j\}_{j\neq 1}} \left(\mathcal{P}^{(1)}_{\boldsymbol{X}_1^{t,\perp}} - \mathcal{P}^{(1)}_{\boldsymbol{U}_1^{\perp}} \right) \right)(\mathcal{Z}) \right\|_{\mathsf{F}} \\
&\leq \left\| \mathcal{M}_1 \left(\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t, \{\boldsymbol{X}_j^t\}_{j\neq 1}} - \mathcal{P}^{(j\neq 1)}_{\mathcal{S}, \{\boldsymbol{U}_j\}_{j\neq 1}} \right)(\mathcal{Z}) \right) \right\|_{\mathsf{F}} + \left\| \boldsymbol{X}_1^t \boldsymbol{X}_1^{t\top} - \boldsymbol{U}_1 \boldsymbol{U}_1^{\top} \right\| \cdot \|\mathcal{Z}\|_{\mathsf{F}} \\
&\leq \left\| \mathcal{M}_1 \left(\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t, \{\boldsymbol{X}_j^t\}_{j\neq 1}} - \mathcal{P}^{(j\neq 1)}_{\mathcal{S}, \{\boldsymbol{U}_j\}_{j\neq 1}} \right)(\mathcal{Z}) \right) \right\|_{\mathsf{F}} + \frac{\left\| \mathcal{X}^t - \mathcal{T} \right\|_{\mathsf{F}}}{\sigma_{\min}(\mathcal{T})} \|\mathcal{Z}\|_{\mathsf{F}}, \qquad (100)
\end{aligned}$$

where we have used the first inequality in this lemma. Denote by $\boldsymbol{Y}_1^t$ and $\boldsymbol{V}_1$ the top-r right singular vectors of $\mathcal{M}_1\left(\mathcal{X}^t\right)$ and $\mathcal{M}_1\left(\mathcal{T}\right)$, respectively. Claim A.1 gives that

$$\begin{aligned}
&\left\| \mathcal{M}_1 \left(\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t, \{\boldsymbol{X}_j^t\}_{j\neq 1}} - \mathcal{P}^{(j\neq 1)}_{\mathcal{S}, \{\boldsymbol{U}_j\}_{j\neq 1}} \right)(\mathcal{Z}) \right) \right\|_{\mathsf{F}} \\
&= \left\| \mathcal{M}_1(\mathcal{Z}) \left(\boldsymbol{Y}_1^t \boldsymbol{Y}_1^{t\top} - \boldsymbol{V}_1 \boldsymbol{V}_1^{\top} \right) \right\|_{\mathsf{F}} \leq \left\| \boldsymbol{Y}_1^t \boldsymbol{Y}_1^{t\top} - \boldsymbol{V}_1 \boldsymbol{V}_1^{\top} \right\| \|\mathcal{Z}\|_{\mathsf{F}} \leq \frac{\left\| \mathcal{X}^t - \mathcal{T} \right\|_{\mathsf{F}}}{\sigma_{\min}(\mathcal{T})} \|\mathcal{Z}\|_{\mathsf{F}}, \quad (101)
\end{aligned}$$

where the last inequality is due to Lemma 33. Plugging (101) into (100) gives

$$\varsigma_1 \leq \frac{2\left\| \mathcal{X}^t - \mathcal{T} \right\|_{\mathsf{F}}}{\sigma_{\min}(\mathcal{T})} \|\mathcal{Z}\|_{\mathsf{F}}.$$

Bounding ς_2 and ς_3. Following a similar argument as above, we can obtain

$$\varsigma_2 \leq \frac{2\left\| \mathcal{X}^t - \mathcal{T} \right\|_{\mathsf{F}}}{\sigma_{\min}(\mathcal{T})} \|\mathcal{Z}\|_{\mathsf{F}} \quad \text{and} \quad \varsigma_3 \leq \frac{2\left\| \mathcal{X}^t - \mathcal{T} \right\|_{\mathsf{F}}}{\sigma_{\min}(\mathcal{T})} \|\mathcal{Z}\|_{\mathsf{F}}.$$

Bounding ς_4. A simple calculation yields that

$$\begin{aligned}
\varsigma_4 &\leq \left\| \mathcal{Z} \times_1 \left(\boldsymbol{X}_1^t \boldsymbol{X}_1^{t\top} - \boldsymbol{U}_1 \boldsymbol{U}_1^{\top} \right) \times_2 \boldsymbol{X}_2^t \boldsymbol{X}_2^{t\top} \times_3 \boldsymbol{X}_3^t \boldsymbol{X}_3^{t\top} \right\|_{\mathsf{F}} \\
&\quad + \left\| \mathcal{Z} \times_1 \boldsymbol{U}_1 \boldsymbol{U}_1^{\top} \times_2 \left(\boldsymbol{X}_2^t \boldsymbol{X}_2^{t\top} - \boldsymbol{U}_2 \boldsymbol{U}_2^{\top} \right) \times_3 \boldsymbol{X}_3^t \boldsymbol{X}_3^{t\top} \right\|_{\mathsf{F}} \\
&\quad + \left\| \mathcal{Z} \times_1 \boldsymbol{U}_1 \boldsymbol{U}_1^{\top} \times_2 \boldsymbol{U}_2 \boldsymbol{U}_2^{\top} \times_3 \left(\boldsymbol{X}_3^t \boldsymbol{X}_3^{t\top} - \boldsymbol{U}_3 \boldsymbol{U}_3^{\top} \right) \right\|_{\mathsf{F}} \\
&\leq 3 \max_{i=1,2,3} \left\| \boldsymbol{X}_i^t \boldsymbol{X}_i^{t\top} - \boldsymbol{U}_i \boldsymbol{U}_i^{\top} \right\| \cdot \|\mathcal{Z}\|_{\mathsf{F}} \\
&\overset{(a)}{\leq} 3 \max_{i=1,2,3} \frac{\left\| \mathcal{M}_i\left(\mathcal{X}^t - \mathcal{T} \right) \right\|_{\mathsf{F}}}{\sigma_{\min}\left(\mathcal{M}_i\left(\mathcal{T} \right) \right)} \|\mathcal{Z}\|_{\mathsf{F}} \leq 3 \frac{\left\| \mathcal{X}^t - \mathcal{T} \right\|_{\mathsf{F}}}{\sigma_{\min}\left(\mathcal{T} \right)} \|\mathcal{Z}\|_{\mathsf{F}},
\end{aligned}$$

where (a) is due to Lemma 33.

Putting together all of the preceding bounds on ς_1, ς_2, ς_3 and ς_4 immediately establishes the lemma.

B.6 Proof of Lemma 15

For any $\mathcal{Z} \in \mathbb{R}^{n\times n\times n}$, we have

$$\begin{aligned}\|\mathcal{P}_\Omega\mathcal{P}_T(\mathcal{Z})\|_{\mathsf{F}}^2 &= \langle \mathcal{P}_\Omega\mathcal{P}_T(\mathcal{Z}), \mathcal{P}_\Omega\mathcal{P}_T(\mathcal{Z})\rangle \\ &= \langle \mathcal{P}_T(\mathcal{Z}), \mathcal{P}_\Omega\mathcal{P}_T(\mathcal{Z})\rangle - p\,\langle \mathcal{P}_T(\mathcal{Z}), \mathcal{P}_T(\mathcal{Z})\rangle + p\,\langle \mathcal{P}_T(\mathcal{Z}), \mathcal{P}_T(\mathcal{Z})\rangle \\ &\le p\left(\left\|\mathcal{P}_T - p^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T\right\| + 1\right)\cdot \|\mathcal{Z}\|_{\mathsf{F}}^2 \le p(1+\varepsilon)\,\|\mathcal{Z}\|_{\mathsf{F}}^2,\end{aligned}$$

which implies that $\|\mathcal{P}_\Omega\mathcal{P}_T\| \le \sqrt{p(1+\varepsilon)}$. Consequently,

$$\begin{aligned}\left\|\mathcal{P}_\Omega\mathcal{P}_{T_{\mathcal{X}^t}}\right\| &\le \left\|\mathcal{P}_\Omega\left(\mathcal{P}_T - \mathcal{P}_{T_{\mathcal{X}^t}}\right)\right\| + \|\mathcal{P}_\Omega\mathcal{P}_T\| \le \left\|\mathcal{P}_T - \mathcal{P}_{T_{\mathcal{X}^t}}\right\| + \|\mathcal{P}_\Omega\mathcal{P}_T\| \\ &\overset{(a)}{\le} \frac{9}{\sigma_{\min}(\mathcal{T})}\left\|\mathcal{X}^t - \mathcal{T}\right\|_{\mathsf{F}} + \sqrt{(1+\epsilon)p} \overset{(b)}{\le} \sqrt{p}\left(\frac{\varepsilon}{1+\varepsilon} + \sqrt{1+\varepsilon}\right) \le 2\sqrt{p}(1+\varepsilon),\end{aligned}$$

where (a) follows from Lemma 14 and (b) is due to the assumption (22).

The spectral norm of $\mathcal{P}_{T_{\mathcal{X}^t}} - p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_\Omega\mathcal{P}_{T_{\mathcal{X}^t}}$ can be bounded as follows:

$$\begin{aligned}\left\|\mathcal{P}_{T_{\mathcal{X}^t}} - p^{-1}\mathcal{P}_{T_{\mathcal{X}^t}}\mathcal{P}_\Omega\mathcal{P}_{T_{\mathcal{X}^t}}\right\| &\le \left\|\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_T\right\| + p^{-1}\left\|\left(\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_T\right)\mathcal{P}_\Omega\mathcal{P}_{T_{\mathcal{X}^t}}\right\| \\ &\quad + p^{-1}\left\|\mathcal{P}_T\mathcal{P}_\Omega\left(\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_T\right)\right\| + \left\|\mathcal{P}_T - p^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T\right\| \\ &\le \left\|\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_T\right\| + p^{-1}\left\|\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_T\right\|\cdot\left\|\mathcal{P}_\Omega\mathcal{P}_{T_{\mathcal{X}^t}}\right\| \\ &\quad + p^{-1}\|\mathcal{P}_T\mathcal{P}_\Omega\|\cdot\left\|\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_T\right\| + \left\|\mathcal{P}_T - p^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T\right\| \\ &\le \left\|\mathcal{P}_{T_{\mathcal{X}^t}} - \mathcal{P}_T\right\|\left(1 + p^{-1}\left\|\mathcal{P}_\Omega\mathcal{P}_{T_{\mathcal{X}^t}}\right\| + p^{-1}\|\mathcal{P}_T\mathcal{P}_\Omega\|\right) + \varepsilon \\ &\le \frac{9}{\sigma_{\min}(\mathcal{T})}\left\|\mathcal{X}^t - \mathcal{T}\right\|_{\mathsf{F}}\left(1 + p^{-1}\frac{\sqrt{p}\varepsilon}{1+\varepsilon} + \frac{2}{p}\|\mathcal{P}_T\mathcal{P}_\Omega\|\right) + \varepsilon \\ &\le \sqrt{p}\frac{\varepsilon}{1+\varepsilon}\left(1 + p^{-1}\frac{\sqrt{p}\varepsilon}{1+\varepsilon} + \frac{2}{p}\cdot\sqrt{p}\cdot\sqrt{1+\varepsilon}\right) + \varepsilon \\ &\le \frac{\varepsilon}{1+\varepsilon}\left(1 + \frac{\varepsilon}{1+\varepsilon} + 2\sqrt{1+\varepsilon}\right) + \varepsilon \\ &\le \frac{\varepsilon}{1+\varepsilon}\left(\frac{1+2\varepsilon+2(1+\varepsilon)^2}{1+\varepsilon}\right) + \varepsilon \le 5\varepsilon,\end{aligned}$$

which completes the proof.

B.7 Proof of Lemma 16

Lemma 16 is similar to Lemma 6 in Cai et al., 2021a, but with three slices being left out instead of one slice. As being noted in Cai et al., 2021b, the proof can be easily adapted from the one slice case to the three slices case. Here we only point out that the structure of Tucker decomposition can be used to simplify the proof in our problem. First the triangle inequality gives that

$$\begin{aligned}&\left\|\mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i^{\ell}\widehat{\boldsymbol{T}}_i^{\ell\mathsf{T}}\right) - \boldsymbol{T}_i\boldsymbol{T}_i^{\mathsf{T}}\right\| \\ &\quad\le \left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}_i^{-1,\ell}\boldsymbol{E}_i^{-1,\ell\mathsf{T}}\right)\right\| + \left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{T}_i\boldsymbol{E}_i^{-1,\ell\mathsf{T}} + \boldsymbol{E}_i^{-1,\ell}\boldsymbol{T}_i^{\mathsf{T}}\right)\right\| + \left\|\mathcal{P}_{\text{diag}}\left(\boldsymbol{T}_i\boldsymbol{T}_i^{\mathsf{T}}\right)\right\|,\end{aligned} \tag{102}$$

where $\boldsymbol{E}_i^{-1,\ell}$ is defined as $\boldsymbol{E}_i^{-1,\ell} = \mathcal{M}_i\left(\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell - \mathcal{I}\right)(\mathcal{T})\right)$.

Bounding $\left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}_i^{-1,\ell}\boldsymbol{E}_i^{-1,\ell\top}\right)\right\|$. By adapting the analysis for $\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}_i^{-1}\boldsymbol{E}_i^{-1}\right)$ in Cai et al., 2021a, Lemma 1 where $\boldsymbol{E}_i^{-1} = \mathcal{M}_i\left(p^{-1}\mathcal{P}_\Omega - \mathcal{I}\right)(\mathcal{T})$, we have

$$\left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}_i^{-1,\ell}\boldsymbol{E}_i^{-1,\ell\top}\right)\right\| \leq C\left(\frac{\mu^{3/2}r^{3/2}}{n^{3/2}p} + \frac{\mu^2 r^2}{n^2 p}\right)\log n \cdot \sigma_{\max}^2(\mathcal{T}). \tag{103}$$

Bounding $\left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{T}_i\boldsymbol{E}_i^{-1,\ell\top} + \boldsymbol{E}_i^{-1,\ell}\boldsymbol{T}_i^\top\right)\right\|$. For this term, we can utilize the Tucker structure to immediately obtain the upper bound. Invoking the triangular inequality gives that

$$\left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{T}_i\boldsymbol{E}_i^{-1,\ell\top} + \boldsymbol{E}_i^{-1,\ell}\boldsymbol{T}_i^\top\right)\right\| \leq \left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{T}_i\boldsymbol{E}_i^{-1,\ell\top}\right)\right\| + \left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}_i^{-1,\ell}\boldsymbol{T}_i^\top\right)\right\|.$$

It suffices to control the spectral norm of $\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{T}_1\boldsymbol{E}_1^{-1,\ell\top}\right)$. To this end, we have

$$\begin{aligned}
\left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{T}_1\boldsymbol{E}_1^{-1,\ell\top}\right)\right\| &\leq 2\left\|\boldsymbol{T}_1\boldsymbol{E}_1^{-1,\ell\top}\right\| = 2\left\|\mathcal{M}_1(\mathcal{T})\mathcal{M}_1^\top(\mathcal{E}^{-1,\ell})\right\| \\
&= 2\left\|\boldsymbol{U}_1\mathcal{M}_1(\mathcal{S})(\boldsymbol{U}_3\otimes\boldsymbol{U}_2)^\top\mathcal{M}_1^\top(\mathcal{E}^{-1,\ell})\right\| \\
&\leq 2\sigma_{\max}(\mathcal{T})\left\|\mathcal{M}_1\left(\mathcal{E}^{-1,\ell}\times_2\boldsymbol{U}_2^\top\times_3\boldsymbol{U}_3^\top\right)\right\| \\
&\overset{(a)}{\leq} 2\sigma_{\max}(\mathcal{T})\sqrt{r}\left\|\mathcal{E}^{-1,\ell}\times_2\boldsymbol{U}_2^\top\times_3\boldsymbol{U}_3^\top\right\| \\
&\leq 2\sigma_{\max}(\mathcal{T})\sqrt{r}\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{T})\right\| \\
&\overset{(b)}{\leq} \sigma_{\max}(\mathcal{T})\cdot\sqrt{r}C\left(\frac{\log^3 n}{p}\|\mathcal{T}\|_\infty + \sqrt{\frac{3\log^5 n}{p}\|\boldsymbol{T}_1^\top\|_{2,\infty}^2}\right),
\end{aligned}$$

where (a) follows from Lemma 22 and (b) is due to Lemma 23. Therefore, the following inequality holds with high probability,

$$\left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{T}_i\boldsymbol{E}_i^{-1,\ell\top} + \boldsymbol{E}_i^{-1,\ell}\boldsymbol{T}_i^\top\right)\right\| \leq C\sigma_{\max}^2(\mathcal{T})\cdot\sqrt{r}\left(\frac{\mu^{3/2}r^{3/2}\log^3 n}{n^{3/2}p} + \sqrt{\frac{\mu^2r^2\log^5 n}{n^2p}}\right). \tag{104}$$

Bounding $\left\|\mathcal{P}_{\text{diag}}\left(\boldsymbol{T}_i\boldsymbol{T}_i^\top\right)\right\|$. A straightforward computation implies that

$$\left\|\mathcal{P}_{\text{diag}}\left(\boldsymbol{T}_i\boldsymbol{T}_i^\top\right)\right\| = \max_{j\in[n]}\left|\left(\boldsymbol{T}_i\boldsymbol{T}_i^\top\right)_{j,j}\right| = \|\boldsymbol{T}_i\|_{2,\infty}^2 \leq \frac{\mu r}{n}\sigma_{\max}^2(\mathcal{T}). \tag{105}$$

Plugging (103), (104) and (105) into (102) yields that, with high probability,

$$\begin{aligned}
&\left\|\mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i^\ell\widehat{\boldsymbol{T}}_i^{\ell\top}\right) - \boldsymbol{T}_i\boldsymbol{T}_i^\top\right\| \\
&\leq C\left(\left(\frac{\mu^{3/2}r^{3/2}}{n^{3/2}p} + \frac{\mu^2r^2}{n^2p}\right)\log n + \sqrt{r}\left(\frac{\mu^{3/2}r^{3/2}\log^3 n}{n^{3/2}p} + \sqrt{\frac{\mu^2r^2\log^5 n}{n^2p}}\right) + \frac{\mu r}{n}\right)\sigma_{\max}^2(\mathcal{T}).
\end{aligned}$$

B.8 Proof of Lemma 17

Recall that $\boldsymbol{X}_i^1 \boldsymbol{\Sigma}_i^1 \boldsymbol{X}_i^{1\top}$ is the top-r eigenvalue decomposition of

$$\boldsymbol{G} := \mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i \widehat{\boldsymbol{T}}_i^\top\right). \tag{106}$$

We denote by $\boldsymbol{G}^\natural$ the Gram matrix $\mathcal{M}_i(\mathcal{T})\mathcal{M}_i^\top(\mathcal{T})$. Applying Lemma 1 in Cai et al., 2021a yields that, with high probability,

$$\begin{aligned}\left\|\boldsymbol{G}-\boldsymbol{G}^\natural\right\| &\leq C\left(\frac{\mu^{3/2}r^{3/2}\sigma_{\max}^2(\mathcal{T})\log n}{n^{3/2}p}+\sqrt{\frac{\mu^2r^2\sigma_{\max}^4(\mathcal{T})\log n}{n^2p}}+\|\mathcal{M}_i(\mathcal{T})\|_{2,\infty}^2\right)\\ &\overset{(a)}{\leq} \frac{1}{2^7}\frac{1}{2^{20}\kappa^6\mu^2r^4}\sigma_{\min}^2(\mathcal{T}),\end{aligned} \tag{107}$$

where (a) follows from (25) and $n \geq C_0\mu^3r^5\kappa^8$. Since $\mathcal{T}$ is a tensor with multilinear rank $\boldsymbol{r}=(r,r,r)$, we have $\sigma_r\left(\boldsymbol{G}^\natural\right)=\sigma_{\min}^2\left(\boldsymbol{T}_i\right)>0$, $\sigma_{r+1}\left(\boldsymbol{G}^\natural\right)=0$. From (107), one can see that $\left\|\boldsymbol{G}-\boldsymbol{G}^\natural\right\| \leq \frac{1}{4}\sigma_r\left(\boldsymbol{G}^\natural\right)$. Therefore, Lemma 26 is applicable, which gives that

$$\begin{aligned}\left\|\boldsymbol{X}_i^1\boldsymbol{R}_i^1-\boldsymbol{U}_i\right\| &\leq \frac{3}{\sigma_r(\boldsymbol{G}^\natural)}\left\|\boldsymbol{G}-\boldsymbol{G}^\natural\right\| \leq \frac{3}{\sigma_{\min}^2(\mathcal{T})}\left\|\boldsymbol{G}-\boldsymbol{G}^\natural\right\|\\ &\leq \frac{3}{\sigma_{\min}^2(\mathcal{T})}\frac{1}{2^7}\frac{1}{2^{20}\kappa^6\mu^2r^4}\sigma_{\min}^2(\mathcal{T}) \leq \frac{1}{2^5}\frac{1}{2^{20}\kappa^6\mu^2r^4}.\end{aligned}$$

Next, we turn to bound $\left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}-\boldsymbol{U}_i\right\|$. Notice that $\boldsymbol{X}_i^{1,\ell}\boldsymbol{\Sigma}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell\top}$ is the top-r eigenvalue decomposition of $\boldsymbol{G}^\ell := \mathcal{P}_{\text{off-diag}}\left(\widehat{\boldsymbol{T}}_i^\ell \widehat{\boldsymbol{T}}_i^{\ell\top}\right)$. By Lemma 16, one can obtain that, with high probability,

$$\begin{aligned}&\left\|\boldsymbol{G}^\ell-\boldsymbol{G}^\natural\right\|\\ &\leq C\left(\left(\frac{\mu^{3/2}r^{3/2}}{n^{3/2}p}+\frac{\mu^2r^2}{n^2p}\right)\log n+\sqrt{r}\left(\frac{\mu^{3/2}r^{3/2}\log^3 n}{n^{3/2}p}+\sqrt{\frac{\mu^2r^2\log^5 n}{n^2p}}\right)+\frac{\mu r}{n}\right)\sigma_{\max}^2\left(\mathcal{T}\right)\\ &\overset{(a)}{\leq} \frac{1}{2^7}\frac{1}{2^{20}\kappa^6\mu^2r^4}\sigma_{\min}^2(\mathcal{T}) \leq \frac{1}{4}\sigma_{\min}^2(\mathcal{T}),\end{aligned}$$

where (a) follows from (25) and $n \geq C_0\mu^3r^5\kappa^8$. Applying Lemma 26 immediately yields that

$$\left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}^{1,\ell}-\boldsymbol{U}_i\right\| \leq \frac{3}{\sigma_{\min}^2(\mathcal{T})}\left\|\boldsymbol{G}^\ell-\boldsymbol{G}^\natural\right\| \leq \frac{3}{\sigma_{\min}^2(\mathcal{T})}\frac{1}{2^7}\frac{1}{2^{20}\kappa^6\mu^2r^4}\sigma_{\min}^2(\mathcal{T}) \leq \frac{1}{2^5}\frac{1}{2^{20}\kappa^6\mu^2r^4}.$$

B.9 Proof of Lemma 18

Noting that $\mathcal{E}^0$ is indeed associated with $\mathcal{X}^1$ (see Equation 14), we need to show Lemma 18 separately for the case $t=1$. According to Lemma 17, one can see that

$$\left\|\boldsymbol{X}_i^1\boldsymbol{R}_i^1-\boldsymbol{U}_i\right\| \leq \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2} \quad \text{and} \quad \left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}-\boldsymbol{U}_i\right\| \leq \frac{1}{2^{25}\kappa^6\mu^2r^4} \leq \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2}.$$

Noticing the definition of $\boldsymbol{G}$ (see Equation 106) and $\boldsymbol{G}^{\natural}$ ($\boldsymbol{G}^{\natural} = \mathcal{M}_i(\mathcal{T})\mathcal{M}_i^{\mathsf{T}}(\mathcal{T})$), the spectral norm of $\boldsymbol{X}_i^1\boldsymbol{X}_i^{1^{\mathsf{T}}} - \boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}$ can be bounded as follows:

$$\begin{aligned}\left\|\boldsymbol{X}_i^1\boldsymbol{X}_i^{1^{\mathsf{T}}} - \boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\right\| &\le \frac{2\left\|\boldsymbol{G}-\boldsymbol{G}^{\natural}\right\|}{\sigma_{\min}^2(\mathcal{T}) - \|\boldsymbol{G}-\boldsymbol{G}^{\natural}\|} \\ &\le \frac{2}{\left(1-\frac{1}{8}\right)\sigma_{\min}^2(\mathcal{T})}\cdot\frac{1}{2^7}\frac{1}{2^{20}\kappa^6\mu^2r^4}\sigma_{\min}^2(\mathcal{T}) \le \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2},\end{aligned}$$

where first line is due to Lemma 34 and Lemma 35, and the second line follows from (107). Using the same argument yields that

$$\left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell^{\mathsf{T}}} - \boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\right\| \le \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2}.$$

By the triangle inequality,

$$\begin{aligned}\left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{X}_i^{1,\ell^{\mathsf{T}}} - \boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\right\|_{2,\infty} &\le \left\|\left(\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}-\boldsymbol{U}_i\right)\left(\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}\right)^{\mathsf{T}}\right\|_{2,\infty} + \left\|\boldsymbol{U}_i\left(\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}-\boldsymbol{U}_i\right)^{\mathsf{T}}\right\|_{2,\infty} \\ &\le \left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}-\boldsymbol{U}_i\right\|_{2,\infty} + \|\boldsymbol{U}_i\|_{2,\infty}\cdot\left\|\boldsymbol{X}_i^{1,\ell}\boldsymbol{R}_i^{1,\ell}-\boldsymbol{U}_i\right\| \\ &\le \frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2}\sqrt{\frac{\mu r}{n}} + \sqrt{\frac{\mu r}{n}}\cdot\frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2} \le \frac{1}{2^{16}\kappa^2\mu^2r^4}\frac{1}{2}\sqrt{\frac{\mu r}{n}}.\end{aligned}$$

For the case $t \ge 2$, first recall that the columns of $\boldsymbol{X}_i^t$ are the top-r eigenvectors of

$$(\boldsymbol{T}_i+\boldsymbol{E}_i^{t-1})(\boldsymbol{T}_i+\boldsymbol{E}_i^{t-1})^{\mathsf{T}} = \boldsymbol{T}_i\boldsymbol{T}_i^{\mathsf{T}} + \underbrace{\boldsymbol{T}_i\boldsymbol{E}_i^{t-1^{\mathsf{T}}} + \boldsymbol{E}_i^{t-1}\boldsymbol{T}_i^{\mathsf{T}} + \boldsymbol{E}_i^{t-1}\boldsymbol{E}_i^{t-1^{\mathsf{T}}}}_{=:\boldsymbol{\Delta}_i^{t-1}}.$$

Thus a simple computation yields that

$$\begin{aligned}\left\|\boldsymbol{\Delta}_i^{t-1}\right\| &\le \left(2\|\boldsymbol{T}_i\| + \left\|\boldsymbol{E}_i^{t-1}\right\|\right)\cdot\left\|\boldsymbol{E}_i^{t-1}\right\| \\ &\overset{(a)}{\le} \left(2\sigma_{\max}(\mathcal{T}) + \frac{1}{8}\sigma_{\max}(\mathcal{T})\right)\cdot\frac{1}{2^{20}\kappa^6\mu^2r^4}\frac{1}{2^t}\sigma_{\max}(\mathcal{T}) \\ &= \frac{17}{2^3}\cdot\frac{1}{2^{20}\kappa^6\mu^2r^4}\frac{1}{2^t}\sigma_{\max}^2(\mathcal{T}) \le \frac{1}{4}\sigma_{\min}^2(\mathcal{T}),\end{aligned}\tag{108}$$

where (a) follows from the induction inequality (17a). Applying Lemma 26 gives that

$$\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{U}_i\right\| \le \frac{3}{\sigma_r(\boldsymbol{T}_i\boldsymbol{T}_i^{\mathsf{T}})}\left\|\boldsymbol{\Delta}_i^{t-1}\right\| \overset{(a)}{\le} \frac{3}{\sigma_{\min}^2(\mathcal{T})}\cdot\frac{17}{2^{23}\kappa^6\mu^2r^4}\frac{1}{2^t}\sigma_{\max}^2(\mathcal{T}) \le \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2^t},$$

where (a) follows from (108). For (43), a direct calculation yields that

$$\begin{aligned}\left\|\boldsymbol{X}_i^t\boldsymbol{X}_i^{t^{\mathsf{T}}} - \boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\right\| &\overset{(a)}{\le} 2\cdot\frac{8}{7\sigma_{\min}^2(\mathcal{T})}\left\|\boldsymbol{\Delta}_i^{t-1}\right\| \\ &\le \frac{16}{7\sigma_{\min}^2(\mathcal{T})}\cdot\frac{17}{2^{23}\kappa^6\mu^2r^4}\frac{1}{2^t}\sigma_{\max}^2(\mathcal{T}) \le \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2^t},\end{aligned}$$

where (a) is due to Lemma 34 and Lemma 35. Following a similar argument, one also has

$$\left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell^{\mathsf{T}}} - \boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\right\| \le \frac{1}{2^{17}\kappa^4\mu^2r^4}\frac{1}{2^t} \text{ and } \left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell^{\mathsf{T}}} - \boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\right\|_{2,\infty} \le \frac{1}{2^{16}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}.$$

B.10 Proof of Lemma 19

Recognizing that $\left\|\boldsymbol{E}_i^{t-1}\right\| \leq \frac{1}{2^{20}\kappa^6\mu^2r^4}\frac{1}{2^t}\sigma_{\max}(\mathcal{T}) \leq \sigma_{\max}(\mathcal{T})/\left(10\kappa^2\right)$ ((17a)), we can apply Lemma 10 to prove this lemma. Combining (17a) and (17f) together gives that

$$B = \max_{i=1,2,3}\left(\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{U}_i\right\|_{2,\infty} + \frac{15\sigma_{\max}(\mathcal{T})}{2\sigma_{\min}^2(\mathcal{T})}\left\|\boldsymbol{U}_i\right\|_{2,\infty}\cdot\left\|\boldsymbol{E}_i^{t-1}\right\|\right)$$
$$\leq \frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}} + \frac{15\sigma_{\max}(\mathcal{T})}{2\sigma_{\min}^2(\mathcal{T})}\cdot\sqrt{\frac{\mu r}{n}}\cdot\frac{1}{2^{20}\kappa^6\mu^2r^4}\frac{1}{2^t}\sigma_{\max}(\mathcal{T}) \leq \frac{9}{2^{20}\kappa^2\mu^2r^4}\cdot\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}.$$

Using the same argument as proving Theorem 6 yields that

$$\left\|\mathcal{X}^t - \mathcal{T}\right\|_\infty \leq \frac{36}{2^{20}\kappa^2\mu^2r^4}\cdot\frac{1}{2^t}\cdot\left(\frac{\mu r}{n}\right)^{3/2}\cdot\sigma_{\max}(\mathcal{T}).$$

Similarly, we have

$$\left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\infty \leq \frac{36}{2^{20}\kappa^2\mu^2r^4}\cdot\frac{1}{2^t}\cdot\left(\frac{\mu r}{n}\right)^{3/2}\cdot\sigma_{\max}(\mathcal{T}).$$

B.11 Proof of Lemma 20

We first prove the inequality (48). Recall that the tensor $\mathcal{X}^t$ can be rewritten as

$$\mathcal{X}^t = \left(\mathcal{T} + \mathcal{E}^{t-1}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top} = \mathcal{T} + \left(\mathcal{T} + \mathcal{E}^{t-1}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top} - \mathcal{T}.$$

Applying the Weyl's inequality reveals that

$$\left|\sigma_k\left(\mathcal{M}_i\left(\mathcal{X}^t\right)\right) - \sigma_k\left(\mathcal{M}_i\left(\mathcal{T}\right)\right)\right| \leq \left\|\mathcal{M}_i\left(\left(\mathcal{T} + \mathcal{E}^{t-1}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top} - \mathcal{T}\right)\right\|$$
$$\leq \underbrace{\left\|\mathcal{M}_i\left(\mathcal{T}\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top} - \mathcal{T}\right)\right\|}_{=:\vartheta_1} + \underbrace{\left\|\mathcal{M}_i\left(\mathcal{E}^{t-1}\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top}\right)\right\|}_{=:\vartheta_2}.$$

Bounding ϑ_1. Using the same argument as controlling (29), one can obtain

$$\vartheta_1 \leq 3\sigma_{\max}(\mathcal{T})\cdot\frac{1}{2^{17}\kappa^4\mu^2r^4} \leq \frac{1}{2^{10}}\sigma_{\min}(\mathcal{T}). \tag{109}$$

Bounding ϑ_2. It is easily seen that

$$\vartheta_2 \leq \max_{i=1,2,3}\left\|\mathcal{M}_i\left(\mathcal{E}^{t-1}\right)\right\| \overset{(a)}{\leq} \frac{1}{2^{20}\kappa^6\mu^2r^4}\frac{1}{2^t}\cdot\sigma_{\max}(\mathcal{T}) \leq \frac{1}{2^{10}}\sigma_{\min}(\mathcal{T}), \tag{110}$$

where (a) is due to (17a) and the fact $\kappa \geq 1$.

Combining (109) and (110) together shows that

$$\left|\sigma_k\left(\mathcal{M}_i\left(\mathcal{X}^t\right)\right) - \sigma_k\left(\mathcal{M}_i\left(\mathcal{T}\right)\right)\right| \leq \frac{1}{2^9}\sigma_{\min}(\mathcal{T}),$$

which implies that

$$\left(1+\frac{1}{2^9}\right)\sigma_{\max}(\mathcal{T}) \geq \sigma_{\max}\left(\mathcal{M}_i\left(\mathcal{X}^t\right)\right) \geq \sigma_{\min}\left(\mathcal{M}_i\left(\mathcal{X}^t\right)\right) \geq \left(1-\frac{1}{2^9}\right)\sigma_{\min}(\mathcal{T}),$$

as well as $\kappa\left(\mathcal{X}^t\right) \leq 2\kappa$. The inequality (49) can be proved in a similar way and the details are omitted.

Appendix C. Proofs of Claims

C.1 Proof of Claim A.1

By the definition of projectors, it is easy to verify that

$$\mathcal{M}_i\left(\left(\mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\perp}}\mathcal{P}^{(j\neq i)}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq i}}\right)(\mathcal{Z})\right) = \mathcal{M}_i\left(\left(\mathcal{P}^{(j\neq i)}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq i}}\mathcal{P}^{(i)}_{\boldsymbol{X}_i^{t,\perp}}\right)(\mathcal{Z})\right),$$

which implies that the above two projectors are commutable. Using the same strategy, one can obtain the second equality.

For the third equality, by symmetry, it suffices to prove $\mathcal{M}_1\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq 1}}(\mathcal{Z})\right) = \mathcal{M}_1(\mathcal{Z})\,\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\top}$. Let $\boldsymbol{Q}_1^t\boldsymbol{\Sigma}_1^t\boldsymbol{Y}_1^{t\top}$ be the singular vector decomposition of $\mathcal{M}_1\left(\mathcal{G}^t\right)\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)^\top$, where $\boldsymbol{Q}_1^t\in\mathbb{R}^{r\times r}$, $\boldsymbol{\Sigma}_1^t\in\mathbb{R}^{r\times r}$ and $\boldsymbol{Y}_1^t\in\mathbb{R}^{n^2\times r}$. We find that

$$\mathcal{M}_1\left(\mathcal{X}^t\right) = \boldsymbol{X}_1^t\mathcal{M}_1\left(\mathcal{G}^t\right)\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)^\top = \boldsymbol{X}_1^t\boldsymbol{Q}_1^t\boldsymbol{\Sigma}_1^t\boldsymbol{Y}_1^{t\top},$$

which implies that the columns of $\boldsymbol{Y}_1^t$ are the top-r right singular vectors of $\mathcal{M}_1\left(\mathcal{X}^t\right)$. A direct calculation gives that

$$\begin{aligned}
\mathcal{M}_1\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq 1}}(\mathcal{Z})\right) &= \mathcal{M}_1(\mathcal{Z})\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)\mathcal{M}_1^\dagger\left(\mathcal{G}^t\right)\mathcal{M}_1\left(\mathcal{G}^t\right)\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)^\top \\
&\overset{(a)}{=} \mathcal{M}_1(\mathcal{Z})\left(\mathcal{M}_1\left(\mathcal{G}^t\right)\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)^\top\right)^\dagger\mathcal{M}_1\left(\mathcal{G}^t\right)\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)^\top \\
&= \mathcal{M}_1(\mathcal{Z})\left(\boldsymbol{Q}_1^t\boldsymbol{\Sigma}_1^t\boldsymbol{Y}_1^{t\top}\right)^\dagger\boldsymbol{Q}_1^t\boldsymbol{\Sigma}_1^t\boldsymbol{Y}_1^{t\top} \\
&= \mathcal{M}_1(\mathcal{Z})\,\boldsymbol{Y}_1^t\boldsymbol{\Sigma}_1^{t\,-1}\boldsymbol{Q}_1^{t\top}\boldsymbol{Q}_1^t\boldsymbol{\Sigma}_1^t\boldsymbol{Y}_1^{t\top} = \mathcal{M}_1(\mathcal{Z})\,\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\top},
\end{aligned} \tag{111}$$

where (a) follows from the fact that $\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t$ has orthonormal columns.

In addition, one has

$$\begin{aligned}
&\mathcal{M}_1\left(\mathcal{P}^{(j\neq 1)}_{\mathcal{G}^t,\left\{\boldsymbol{X}_j^t\right\}_{j\neq 1}}(\mathcal{Z})\right) \\
&= \mathcal{M}_1(\mathcal{Z})\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)^\top\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)\mathcal{M}_1^\dagger\left(\mathcal{G}^t\right)\mathcal{M}_1\left(\mathcal{G}^t\right)\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)^\top \\
&= \mathcal{M}_1(\mathcal{Z})\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)\left(\boldsymbol{X}_3^t\otimes\boldsymbol{X}_2^t\right)^\top\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\top}.
\end{aligned} \tag{112}$$

The last two inequalities in Claim A.1 had already been used in Cai et al., 2020and can be easily verified by definition. Thus we omit the proof here.

C.2 Proof of Claim A.2

By the equation (53), we have

$$
\begin{aligned}
\beta_{3,a} &= \left\| \boldsymbol{e}_m^\top \mathcal{M}_1 \left(\left(\mathcal{I} - \mathcal{P}_{T_{\mathcal{X}^{t,\ell}}} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2 \\
&\le \sum_{i=1}^{3} \underbrace{\left\| \boldsymbol{e}_m^\top \mathcal{M}_1 \left(\left(\mathcal{P}_{\boldsymbol{U}_i}^{(i)} - \mathcal{P}_{\boldsymbol{X}_i^{t,\ell}}^{(i)} \right) \left(\prod_{j\neq i} \mathcal{P}_{\boldsymbol{X}_j^{t,\ell}}^{(j)} - \mathcal{P}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j\neq i}}^{(j\neq i)} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2}_{=:\beta_{3,a}^{i}} \\
&\quad + \underbrace{\left\| \boldsymbol{e}_m^\top \mathcal{M}_1 \left(\left(\mathcal{P}_{\boldsymbol{U}_1}^{(1)} - \mathcal{P}_{\boldsymbol{X}_1^{t,\ell}}^{(1)} \right) \mathcal{P}_{\boldsymbol{X}_2^{t,\ell,\perp}}^{(2)} \mathcal{P}_{\boldsymbol{X}_3^{t,\ell}}^{(3)} \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2}_{=:\beta_{3,a}^{4}} \\
&\quad + \underbrace{\left\| \boldsymbol{e}_m^\top \mathcal{M}_1 \left(\left(\mathcal{P}_{\boldsymbol{U}_3}^{(3)} - \mathcal{P}_{\boldsymbol{X}_3^{t,\ell}}^{(3)} \right) \mathcal{P}_{\boldsymbol{X}_1^{t,\ell,\perp}}^{(1)} \mathcal{P}_{\boldsymbol{X}_2^{t,\ell}}^{(2)} \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2}_{=:\beta_{3,a}^{5}} \\
&\quad + \underbrace{\left\| \boldsymbol{e}_m^\top \mathcal{M}_1 \left(\left(\mathcal{P}_{\boldsymbol{U}_2}^{(2)} - \mathcal{P}_{\boldsymbol{X}_2^{t,\ell}}^{(2)} \right) \mathcal{P}_{\boldsymbol{X}_3^{t,\ell,\perp}}^{(3)} \mathcal{P}_{\boldsymbol{X}_1^{t,\ell}}^{(1)} \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2}_{=:\beta_{3,a}^{6}} \\
&\quad + \underbrace{\left\| \boldsymbol{e}_m^\top \mathcal{M}_1 \left(\left(\mathcal{P}_{\boldsymbol{U}_1}^{(1)} - \mathcal{P}_{\boldsymbol{X}_1^{t,\ell}}^{(1)} \right) \mathcal{P}_{\boldsymbol{X}_2^{t,\ell,\perp}}^{(2)} \mathcal{P}_{\boldsymbol{X}_3^{t,\ell,\perp}}^{(3)} \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2}_{=:\beta_{3,a}^{7}} .
\end{aligned}
$$

Bounding $\beta_{3,a}^1$. A straightforward computation yields that

$$
\begin{aligned}
\beta_{3,a}^1 &\le \left\| \boldsymbol{e}_m^\top \left(\boldsymbol{U}_1 \boldsymbol{U}_1^\top - \boldsymbol{X}_1^{t,\ell} \boldsymbol{X}_1^{t,\ell\top} \right) \right\|_2 \cdot \left\| \prod_{j\neq 1} \mathcal{P}_{\boldsymbol{X}_j^{t,\ell}}^{(j)} - \mathcal{P}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j\neq 1}}^{(j\neq 1)} \right\| \cdot \left\| \mathcal{X}^{t,\ell} - \mathcal{T} \right\|_{\mathsf{F}} \\
&= \left\| \boldsymbol{e}_m^\top \left(\boldsymbol{U}_1 \boldsymbol{U}_1^\top - \boldsymbol{X}_1^{t,\ell} \boldsymbol{X}_1^{t,\ell\top} \right) \right\|_2 \left\| \mathcal{X}^{t,\ell} - \mathcal{T} \right\|_{\mathsf{F}} ,
\end{aligned}
$$

where the last line follows from Claim A.1.

Bounding $\beta_{3,a}^2$ and $\beta_{3,a}^3$. It follows from the triangle inequality that

$$
\begin{aligned}
\beta_{3,a}^2 &= \left\| \boldsymbol{e}_m^\top \mathcal{M}_1 \left(\left(\mathcal{P}_{\boldsymbol{U}_2}^{(2)} - \mathcal{P}_{\boldsymbol{X}_2^{t,\ell}}^{(2)} \right) \left(\prod_{j\neq 2} \mathcal{P}_{\boldsymbol{X}_j^{t,\ell}}^{(j)} - \mathcal{P}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j\neq 2}}^{(j\neq 2)} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2 \\
&\le \left\| \boldsymbol{e}_m^\top \mathcal{M}_1 \left(\left(\mathcal{P}_{\boldsymbol{U}_2}^{(2)} - \mathcal{P}_{\boldsymbol{X}_2^{t,\ell}}^{(2)} \right) \left(\prod_{j\neq 2} \mathcal{P}_{\boldsymbol{X}_j^{t,\ell}}^{(j)} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2 \\
&\quad + \left\| \boldsymbol{e}_m^\top \mathcal{M}_1 \left(\left(\mathcal{P}_{\boldsymbol{U}_2}^{(2)} - \mathcal{P}_{\boldsymbol{X}_2^{t,\ell}}^{(2)} \right) \left(\mathcal{P}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j\neq 2}}^{(j\neq 2)} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2 .
\end{aligned}
$$

The first term on the right hand side can be bounded by

$$\begin{aligned}
&\left\| \boldsymbol{e}_m^{\mathsf{T}} \mathcal{M}_1 \left(\left(\mathcal{P}^{(2)}_{\boldsymbol{U}_2} - \mathcal{P}^{(2)}_{\boldsymbol{X}_2^{t,\ell}} \right) \left(\prod_{j \neq 2} \mathcal{P}^{(j)}_{\boldsymbol{X}_j^{t,\ell}} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2 \\
&\quad = \left\| \boldsymbol{e}_m^{\mathsf{T}} \boldsymbol{X}_1^{t,\ell} \boldsymbol{X}_1^{t,\ell\mathsf{T}} \mathcal{M}_1 \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \left(\left(\boldsymbol{X}_3^{t,\ell} \boldsymbol{X}_3^{t,\ell\mathsf{T}} \right) \otimes \left(\boldsymbol{U}_2 \boldsymbol{U}_2^{\mathsf{T}} - \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\mathsf{T}} \right) \right) \right\|_2 \\
&\quad \le \left\| \boldsymbol{e}_m^{\mathsf{T}} \boldsymbol{X}_1^{t,\ell} \right\|_2 \left\| \mathcal{M}_1 \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right\| \cdot \left\| \left(\boldsymbol{X}_3^{t,\ell} \boldsymbol{X}_3^{t,\ell\mathsf{T}} \right) \otimes \left(\boldsymbol{U}_2 \boldsymbol{U}_2^{\mathsf{T}} - \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\mathsf{T}} \right) \right\| \\
&\quad \le \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \left\| \boldsymbol{U}_2 \boldsymbol{U}_2^{\mathsf{T}} - \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\mathsf{T}} \right\| \cdot \left\| \mathcal{X}^{t,\ell} - \mathcal{T} \right\|_{\mathsf{F}}.
\end{aligned}$$

By the definition of $\mathcal{P}^{(j \neq 2)}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j \neq 2}}$ in (51), we have

$$\begin{aligned}
&\mathcal{M}_2 \left(\mathcal{P}^{(j \neq 2)}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j \neq 2}} \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \\
&\quad = \mathcal{M}_2 \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_1^{t,\ell} \right) \mathcal{M}_2^{\dagger} \left(\mathcal{G}^{t,\ell} \right) \mathcal{M}_2 \left(\mathcal{G}^{t,\ell} \right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_1^{t,\ell} \right)^{\mathsf{T}} \\
&\quad = \boldsymbol{W}_2^{t,\ell} \mathcal{M}_2 \left(\mathcal{G}^{t,\ell} \right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_1^{t,\ell} \right)^{\mathsf{T}} = \mathcal{M}_2 \left(\mathcal{G}^{t,\ell} \times_1 \boldsymbol{X}_1^{t,\ell} \times_2 \boldsymbol{W}_2^{t,\ell} \times_3 \boldsymbol{X}_3^{t,\ell} \right),
\end{aligned}$$

where $\boldsymbol{W}_2^{t,\ell} = \mathcal{M}_2(\mathcal{X}^{t,\ell} - \mathcal{T}) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_1^{t,\ell} \right) \mathcal{M}_2^{\dagger} \left(\mathcal{G}^{t,\ell} \right)$. It follows that

$$\mathcal{P}^{(j \neq 2)}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j \neq 2}} \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) = \mathcal{G}^{t,\ell} \times_1 \boldsymbol{X}_1^{t,\ell} \times_2 \boldsymbol{W}_2^{t,\ell} \times_3 \boldsymbol{X}_3^{t,\ell}.$$

Thus one has

$$\begin{aligned}
&\left\| \boldsymbol{e}_m^{\mathsf{T}} \mathcal{M}_1 \left(\left(\mathcal{P}^{(2)}_{\boldsymbol{U}_2} - \mathcal{P}^{(2)}_{\boldsymbol{X}_2^{t,\ell}} \right) \left(\mathcal{P}^{(j \neq 2)}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j \neq 2}} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2 \\
&\quad \le \left\| \boldsymbol{e}_m^{\mathsf{T}} \mathcal{M}_1 \left(\left(\mathcal{P}^{(j \neq 2)}_{\mathcal{G}^{t,\ell}, \left\{ \boldsymbol{X}_j^{t,\ell} \right\}_{j \neq 2}} \right) \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right) \right\|_2 \cdot \left\| \boldsymbol{U}_2 \boldsymbol{U}_2^{\mathsf{T}} - \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\mathsf{T}} \right\| \\
&\quad = \left\| \boldsymbol{e}_m^{\mathsf{T}} \mathcal{M}_1 \left(\mathcal{G}^{t,\ell} \times_1 \boldsymbol{X}_1^{t,\ell} \times_2 \boldsymbol{W}_2^{t,\ell} \times_3 \boldsymbol{X}_3^{t,\ell} \right) \right\|_2 \cdot \left\| \boldsymbol{U}_2 \boldsymbol{U}_2^{\mathsf{T}} - \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\mathsf{T}} \right\| \\
&\quad \le \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \sigma_1 \left(\mathcal{M}_1 \left(\mathcal{X}^{t,\ell} \right) \right) \cdot \left\| \boldsymbol{W}_2^{t,\ell} \right\| \cdot \left\| \boldsymbol{U}_2 \boldsymbol{U}_2^{\mathsf{T}} - \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\mathsf{T}} \right\| \\
&\quad \le \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \kappa \left(\mathcal{X}^{t,\ell} \right) \cdot \left\| \mathcal{X}^{t,\ell} - \mathcal{T} \right\|_{\mathsf{F}} \cdot \left\| \boldsymbol{U}_2 \boldsymbol{U}_2^{\mathsf{T}} - \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\mathsf{T}} \right\| \qquad (113)
\end{aligned}$$

where the last line follows from

$$\left\| \boldsymbol{W}_2^{t,\ell} \right\| \le \left\| \mathcal{M}_2 \left(\mathcal{X}^{t,\ell} - \mathcal{T} \right) \right\| \cdot \left\| \mathcal{M}_2^{\dagger} \left(\mathcal{G}^{t,\ell} \right) \right\| \le \left\| \mathcal{X}^{t,\ell} - \mathcal{T} \right\|_{\mathsf{F}} \cdot \frac{1}{\sigma_r \left(\mathcal{M}_2 \left(\mathcal{X}^{t,\ell} \right) \right)}.$$

Consequently,

$$\beta_{3,a}^2 \le \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \left\| \boldsymbol{U}_2 \boldsymbol{U}_2^{\mathsf{T}} - \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\mathsf{T}} \right\| \cdot \left\| \mathcal{X}^{t,\ell} - \mathcal{T} \right\|_{\mathsf{F}} \cdot \left(1 + \kappa \left(\mathcal{X}^{t,\ell} \right) \right)$$

$$\leq \left\|\boldsymbol{X}_1^{t,\ell}\right\|_{2,\infty} \cdot \left\|\boldsymbol{U}_2\boldsymbol{U}_2^\top - \boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\right\| \cdot \left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\mathsf{F} \cdot (1+2\kappa).$$

Using the same argument, one can obtain

$$\beta_{3,a}^3 \leq \left\|\boldsymbol{X}_1^{t,\ell}\right\|_{2,\infty} \cdot \left\|\boldsymbol{U}_3\boldsymbol{U}_3^\top - \boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\right\| \cdot \left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\mathsf{F} \cdot (1+2\kappa).$$

Bounding $\beta_{3,a}^j$ for $j = 4,5,6,7$. A straightforward computation yields that

$$\begin{aligned}
\beta_{3,a}^4 &\leq \left\|\boldsymbol{e}_m^\top \left(\boldsymbol{U}_1\boldsymbol{U}_1^\top - \boldsymbol{X}_1^{t,\ell}\boldsymbol{X}_1^{t,\ell\top}\right)\right\|_2 \cdot \left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\mathsf{F},\\
\beta_{3,a}^5 &\leq \left\|\boldsymbol{X}_1^{t,\ell}\right\|_{2,\infty} \cdot \left\|\boldsymbol{U}_3\boldsymbol{U}_3^\top - \boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\right\| \cdot \left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\mathsf{F},\\
\beta_{3,a}^6 &\leq \left\|\boldsymbol{X}_1^{t,\ell}\right\|_{2,\infty} \cdot \left\|\boldsymbol{U}_2\boldsymbol{U}_2^\top - \boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\right\| \cdot \left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\mathsf{F},\\
\beta_{3,a}^7 &\leq \left\|\boldsymbol{e}_m^\top \left(\boldsymbol{U}_1\boldsymbol{U}_1^\top - \boldsymbol{X}_1^{t,\ell}\boldsymbol{X}_1^{t,\ell\top}\right)\right\|_2 \cdot \left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\mathsf{F}.
\end{aligned}$$

Combining all the bounds together shows that

$$\begin{aligned}
\beta_{3,a} &\leq \left(3\left\|\boldsymbol{e}_m^\top \left(\boldsymbol{U}_1\boldsymbol{U}_1^\top - \boldsymbol{X}_1^{t,\ell}\boldsymbol{X}_1^{t,\ell\top}\right)\right\|_2 + 8\kappa\left\|\boldsymbol{X}_1^{t,\ell}\right\|_{2,\infty} \max_{i=2,3}\left\|\boldsymbol{U}_i\boldsymbol{U}_i^\top - \boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}\right\|\right)\left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\mathsf{F}\\
&\overset{(a)}{\leq} \left(\frac{3}{2^{16}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}} + \frac{16\kappa}{2^{16}\kappa^4\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\right)\cdot n^{3/2}\cdot\frac{36}{2^{20}\kappa^2\mu^2r^4}\cdot\frac{1}{2^t}\cdot\left(\frac{\mu r}{n}\right)^{3/2}\cdot\sigma_{\max}(\mathcal{T})\\
&\leq \frac{19\cdot 36}{2^{16}}\cdot\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}) \leq \frac{1}{2^6}\cdot\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}),
\end{aligned}$$

where (a) follows from Lemma 18.

C.3 Proof of Claim A.3

Let $\mathcal{Z}^{t,\ell} := \left(\mathcal{I} - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)$. By the definition of $\mathcal{P}_{T_{\mathcal{X}^{t,\ell}}}$ in (8), the term $\beta_{3,b}$ can be bounded as follows:

$$\begin{aligned}
\beta_{3,b} &= \left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top} + \sum_{i=1}^3 \mathcal{G}^{t,\ell}\times_i \boldsymbol{W}_i^{t,\ell}\underset{j\neq i}{\times}\boldsymbol{X}_j^{t,\ell}\right)\right\|_2\\
&\leq \sum_{i=1}^3 \underbrace{\left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{G}^{t,\ell}\times_i\boldsymbol{W}_i^{t,\ell}\underset{j\neq i}{\times}\boldsymbol{X}_j^{t,\ell}\right)\right\|_2}_{=:\beta_{3,b}^i} + \underbrace{\left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\overset{3}{\underset{i=1}{\times}}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}\right)\right\|_2}_{=:\beta_{3,b}^4},
\end{aligned}$$

where $\boldsymbol{W}_i^{t,\ell}$ is given by $\boldsymbol{W}_i^{t,\ell} = \left(\boldsymbol{I} - \boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}\right)\mathcal{M}_i\left(\mathcal{Z}^{t,\ell}\underset{j\neq i}{\times}\boldsymbol{X}_j^{t,\ell\top}\right)\mathcal{M}_i^\dagger\left(\mathcal{G}^{t,\ell}\right)$.

Bounding $\beta_{3,b}^1$. It can be bounded as follows:

$$\beta_{3,b}^1 = \left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{G}^{t,\ell}\times_1\boldsymbol{W}_1^{t,\ell}\times_2\boldsymbol{X}_2^{t,\ell}\times_3\boldsymbol{X}_3^{t,\ell}\right)\right\|_2 = \left\|\boldsymbol{e}_m^\top\boldsymbol{W}_1^{t,\ell}\mathcal{M}_1\left(\mathcal{G}^{t,\ell}\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)^\top\right\|_2$$

$$
\begin{aligned}
&= \left\| \boldsymbol{e}_m^\top \left(\boldsymbol{I} - \boldsymbol{X}_1^{t,\ell} \boldsymbol{X}_1^{t,\ell\top} \right) \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right) \mathcal{M}_1^\dagger\left(\mathcal{G}^{t,\ell}\right) \mathcal{M}_1\left(\mathcal{G}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right)^\top \right\|_2 \\
&\le \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right) \mathcal{M}_1^\dagger\left(\mathcal{G}^{t,\ell}\right) \mathcal{M}_1\left(\mathcal{G}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right)^\top \right\|_2 \\
&\quad + \left\| \boldsymbol{e}_m^\top \boldsymbol{X}_1^{t,\ell} \boldsymbol{X}_1^{t,\ell\top} \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right) \mathcal{M}_1^\dagger\left(\mathcal{G}^{t,\ell}\right) \mathcal{M}_1\left(\mathcal{G}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right)^\top \right\|_2 \\
&\overset{(a)}{\le} \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right) \right\|_2 + \left\| \boldsymbol{e}_m^\top \boldsymbol{X}_1^{t,\ell} \right\|_2 \cdot \left\| \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right) \right\| \\
&= \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right) \right\|_2 + \left\| \boldsymbol{e}_m^\top \boldsymbol{X}_1^{t,\ell} \right\|_2 \cdot \left\| \mathcal{M}_1\left(\mathcal{Z}^{t,\ell} \times_2 \boldsymbol{X}_2^{t,\ell\top} \times_3 \boldsymbol{X}_3^{t,\ell\top} \right) \right\| \\
&\overset{(b)}{\le} \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right) \right\|_2 + \sqrt{r} \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \left\| \mathcal{Z}^{t,\ell} \times_2 \boldsymbol{X}_2^{t,\ell\top} \times_3 \boldsymbol{X}_3^{t,\ell\top} \right\| \\
&\le \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right) \right\|_2 + \sqrt{r} \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \left\| \mathcal{Z}^{t,\ell} \right\|,
\end{aligned}
$$

where step (a) is due to the fact $\left\| \mathcal{M}_1^\dagger\left(\mathcal{G}^{t,\ell}\right) \mathcal{M}_1\left(\mathcal{G}^{t,\ell}\right) \right\| \le 1$ and step (b) has used Lemma 22.

Bounding $\beta_{3,b}^2$ and $\beta_{3,b}^3$. A direct calculation yields that

$$
\begin{aligned}
\beta_{3,b}^2 &= \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{G}^{t,\ell} \times_1 \boldsymbol{X}_1^{t,\ell} \times_2 \boldsymbol{W}_2^{t,\ell} \times_3 \boldsymbol{X}_3^{t,\ell} \right) \right\|_2 \\
&= \left\| \boldsymbol{e}_m^\top \boldsymbol{X}_1^{t,\ell} \mathcal{M}_1\left(\mathcal{G}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{W}_2^{t,\ell} \right)^\top \right\|_2 \le \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \left\| \mathcal{M}_1\left(\mathcal{G}^{t,\ell}\right) \right\| \cdot \left\| \boldsymbol{W}_2^{t,\ell} \right\| \\
&\le \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \left\| \mathcal{M}_1\left(\mathcal{G}^{t,\ell}\right) \right\| \cdot \left\| \mathcal{M}_2\left(\mathcal{Z}^{t,\ell} \underset{j\neq 2}{\times} \boldsymbol{X}_j^{t,\ell\top} \right) \right\| \cdot \left\| \mathcal{M}_2^\dagger\left(\mathcal{G}^{t,\ell}\right) \right\| \\
&\le \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \sqrt{r} \left\| \mathcal{Z}^{t,\ell} \right\| \cdot \kappa\left(\mathcal{X}^{t,\ell}\right) \le 2\kappa\sqrt{r} \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \left\| \mathcal{Z}^{t,\ell} \right\|,
\end{aligned}
$$

where the last line is due to $\kappa\left(\mathcal{X}^{t,\ell}\right) \le 2\kappa$, see Lemma 20. Using the same argument, one can obtain

$$
\beta_{3,b}^3 = \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{G}^{t,\ell} \times_1 \boldsymbol{X}_1^{t,\ell} \times_2 \boldsymbol{X}_2^{t,\ell} \times_3 \boldsymbol{W}_3^{t,\ell} \right) \right\|_2 \le 2\kappa\sqrt{r} \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \left\| \mathcal{Z}^{t,\ell} \right\|.
$$

Bounding $\beta_{3,b}^4$. It can be bounded as follows:

$$
\begin{aligned}
\beta_{3,b}^4 &= \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell} \overset{3}{\underset{i=1}{\times}} \boldsymbol{X}_i^{t,\ell} \boldsymbol{X}_i^{t,\ell\top} \right) \right\|_2 = \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell} \overset{3}{\underset{i=1}{\times}} \boldsymbol{X}_i^{t,\ell\top} \overset{3}{\underset{i=1}{\times}} \boldsymbol{X}_i^{t,\ell} \right) \right\|_2 \\
&= \left\| \boldsymbol{e}_m^\top \boldsymbol{X}_1^{t,\ell} \mathcal{M}_1\left(\mathcal{Z}^{t,\ell} \overset{3}{\underset{i=1}{\times}} \boldsymbol{X}_i^{t,\ell\top} \right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right)^\top \right\|_2 \le \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \left\| \mathcal{M}_1\left(\mathcal{Z}^{t,\ell} \overset{3}{\underset{i=1}{\times}} \boldsymbol{X}_i^{t,\ell\top} \right) \right\| \\
&\overset{(a)}{\le} \sqrt{r} \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \left\| \mathcal{Z}^{t,\ell} \overset{3}{\underset{i=1}{\times}} \boldsymbol{X}_i^{t,\ell\top} \right\| \le \sqrt{r} \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \left\| \mathcal{Z}^{t,\ell} \right\|,
\end{aligned}
$$

where step (a) is due to Lemma 22.

Combining these four terms together reveals that

$$
\beta_{3,b} \le \underbrace{\left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right) \right\|_2}_{=:\gamma} + 6\kappa \left\| \boldsymbol{X}_1^{t,\ell} \right\|_{2,\infty} \cdot \sqrt{r} \left\| \mathcal{Z}^{t,\ell} \right\|. \tag{114}
$$

For the second term on the right hand side, we apply Lemma 25 to show that

$$\begin{aligned}
6\kappa \left\|\boldsymbol{X}_1^{t,\ell}\right\|_{2,\infty} \cdot \sqrt{r}\left\|\mathcal{Z}^{t,\ell}\right\| &= 6\kappa \left\|\boldsymbol{X}_1^{t,\ell}\right\|_{2,\infty} \cdot \sqrt{r}\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega_{-\ell}}-\mathcal{P}_\ell\right)\left(\mathcal{X}^{t,\ell}-\mathcal{T}\right)\right\| \\
&\le 12\kappa\sqrt{\frac{\mu r}{n}}\sqrt{r}\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega}\right)\left(\mathcal{X}^{t,\ell}-\mathcal{T}\right)\right\| \\
&\le 12\kappa\sqrt{\frac{\mu r}{n}}\sqrt{r}\left\|\left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega}\right)(\mathcal{J})\right\| \cdot 2r\left\|\mathcal{X}^{t,\ell}-\mathcal{T}\right\|_\infty \\
&\le \frac{1}{2^9}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \qquad (115)
\end{aligned}$$

where the second line is due to $\left\|\boldsymbol{X}_1^{t,\ell}\right\|_{2,\infty} \le \left\|\boldsymbol{X}_1^{t,\ell}\boldsymbol{R}_1^{t,\ell}-\boldsymbol{U}_1\right\|_{2,\infty}+\|\boldsymbol{U}_1\|_{2,\infty} \le 2\sqrt{\frac{\mu r}{n}}$ and the last step has used the assumption $p \ge \max\left\{\frac{C_1\kappa^3\mu^{1.5}r^3\log^3 n}{n^{3/2}}, \frac{C_2\kappa^6\mu^3r^6\log^5 n}{n^2}\right\}$.

Next, we turn to control the term γ. Recall that $\mathcal{Z}^{t,\ell}=\left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega_{-\ell}}-\mathcal{P}_\ell\right)\left(\mathcal{X}^{t,\ell}-\mathcal{T}\right)$. Letting

$$\begin{aligned}
\mathcal{Z}^{t,m} &:= \left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega_{-\ell}}-\mathcal{P}_\ell\right)\left(\mathcal{X}^{t,m}-\mathcal{T}\right), \\
\mathcal{Z}^{t,m,\ell} &:= \left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega_{-\ell}}-\mathcal{P}_\ell\right)\left(\mathcal{X}^{t,\ell}-\mathcal{X}^{t,m}\right),
\end{aligned}$$

we have $\mathcal{Z}^{t,\ell}=\mathcal{Z}^{t,m}+\mathcal{Z}^{t,m,\ell}$. If $m=\ell$, it is not hard to see that $\gamma=0$. Thus, without loss of generality, we assume $m \ne \ell$. Denoting by $\boldsymbol{O}_i=\boldsymbol{R}_i^{t,m}\boldsymbol{R}_i^{t,\ell\top}$ the special orthogonal matrix, the triangle inequality gives that

$$\begin{aligned}
\gamma &= \left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)\right\|_2 \\
&\le \left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}-\boldsymbol{X}_3^{t,m}\boldsymbol{O}_3\otimes\boldsymbol{X}_2^{t,m}\boldsymbol{O}_2\right)\right\|_2+\left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right)\left(\boldsymbol{X}_3^{t,m}\boldsymbol{O}_3\otimes\boldsymbol{X}_2^{t,m}\boldsymbol{O}_2\right)\right\|_2 \\
&\le \left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}-\boldsymbol{X}_3^{t,m}\boldsymbol{O}_3\otimes\boldsymbol{X}_2^{t,m}\boldsymbol{O}_2\right)\right\|_2+\left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,m}\right)\left(\boldsymbol{X}_3^{t,m}\boldsymbol{O}_3\otimes\boldsymbol{X}_2^{t,m}\boldsymbol{O}_2\right)\right\|_2 \\
&\quad+\left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,m,\ell}\right)\left(\boldsymbol{X}_3^{t,m}\boldsymbol{O}_3\otimes\boldsymbol{X}_2^{t,m}\boldsymbol{O}_2\right)\right\|_2 \\
&= \left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}-\boldsymbol{X}_3^{t,m}\boldsymbol{O}_3\otimes\boldsymbol{X}_2^{t,m}\boldsymbol{O}_2\right)\right\|_2+\left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,m}\right)\left(\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right)\right\|_2 \\
&\quad+\left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,m,\ell}\right)\left(\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right)\right\|_2 \\
&:= \gamma_1+\gamma_2+\gamma_3.
\end{aligned}$$

For the sake of clarity, we first give the upper bounds for the above three terms whose proofs can be found in Sections C.3.1, C.3.2 and C.3.3:

$$\gamma_1 \le \frac{1}{2^{11}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \qquad (116)$$

$$\gamma_2 \le \frac{1}{2^{11}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \qquad (117)$$

$$\gamma_3 \le \frac{1}{2^{11}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}). \qquad (118)$$

Thus we have

$$\gamma \le \gamma_1 + \gamma_2 + \gamma_3 \le \frac{3}{2^{11}} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}). \tag{119}$$

Plugging (119) and (115) into (114) yields that

$$\beta_{3,b} \le \frac{1}{2^6} \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}).$$

C.3.1 Proof of (116)

A simple calculation gives that

$$\begin{aligned}
\gamma_1 &= \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} - \boldsymbol{X}_3^{t,m} \boldsymbol{R}_3^{t,m} \boldsymbol{R}_3^{t,\ell\top} \otimes \boldsymbol{X}_2^{t,m} \boldsymbol{R}_2^{t,m} \boldsymbol{R}_2^{t,\ell\top} \right) \right\|_2 \\
&= \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} - \boldsymbol{X}_3^{t,m} \boldsymbol{R}_3^{t,m} \boldsymbol{R}_3^{t,\ell\top} \otimes \boldsymbol{X}_2^{t,m} \boldsymbol{R}_2^{t,m} \boldsymbol{R}_2^{t,\ell\top} \right) \left(\boldsymbol{R}_3^{t,\ell} \otimes \boldsymbol{R}_2^{t,\ell} \right) \right\|_2 \\
&= \left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} - \boldsymbol{X}_3^{t,m} \boldsymbol{R}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m} \boldsymbol{R}_2^{t,m} \right) \right\|_2 \\
&\le \underbrace{\left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \right) \right\|_2}_{=:\gamma_{1,a}} \\
&\quad + \underbrace{\left\| \boldsymbol{e}_m^\top \mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right) \left(\boldsymbol{X}_3^{t,m} \boldsymbol{R}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m} \boldsymbol{R}_2^{t,m} - \boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \right) \right\|_2}_{=:\gamma_{1,b}}.
\end{aligned}$$

For the first term $\gamma_{1,a}$, it can be bounded as follows:

$$\begin{aligned}
\gamma_{1,a} &= \sqrt{\sum_{s=1}^{r^2} \left(\sum_{j=1}^{n^2} \left[\mathcal{M}_1\left(\mathcal{Z}^{t,\ell}\right)\right]_{m,j} \left[\boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \right]_{j,s} \right)^2} \\
&= \sqrt{\sum_{s=1}^{r^2} \left(\sum_{j\notin\Gamma} \left(p^{-1}\delta_{m,j} - 1\right) \left[\mathcal{M}_1\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right]_{m,j} \left[\boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \right]_{j,s} \right)^2} \\
&\le r \max_{s\in[r^2]} \left| \sum_{j\notin\Gamma} \left(p^{-1}\delta_{m,j} - 1\right) \left[\mathcal{M}_1\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right]_{m,j} \left[\boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \right]_{j,s} \right| \\
&\le r \max_{s\in[r^2]} \sum_{j\notin\Gamma} \left|p^{-1}\delta_{m,j} - 1\right| \cdot \left| \left[\mathcal{M}_1\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right]_{m,j} \right| \cdot \left| \left[\boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \right]_{j,s} \right| \\
&\le r \sum_{j\notin\Gamma} \left(p^{-1}\delta_{m,j} + 1\right) \cdot \left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\infty \cdot \left\| \boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \right\|_\infty \\
&\overset{(a)}{\le} r \cdot 3n^2 \cdot \left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\infty \cdot \left\| \boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \right\|_{2,\infty} \qquad (120) \\
&\le r \cdot 3n^2 \cdot \left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\infty \left(\left\| \boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} - \boldsymbol{U}_3 \right\|_{2,\infty} \left\| \boldsymbol{X}_2^{t,\ell} \right\|_{2,\infty} + \left\| \boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_2 \right\|_{2,\infty} \left\| \boldsymbol{U}_3 \right\|_{2,\infty} \right)
\end{aligned}$$

$$\overset{(b)}{\leq} 3n^2r \cdot \frac{36}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\left(\frac{\mu r}{n}\right)^{3/2}\sigma_{\max}(\mathcal{T}) \cdot \frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}} \cdot \frac{7}{3}\sqrt{\frac{\mu r}{n}}$$
$$\leq \frac{1}{2^{12}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}).$$

Here in (a), we use the fact that whenever $p \geq \frac{C_1 \log n}{n^{3/2}}$, with high probability, all rows have at most $2n^2p$ observed entries (Arratia and Gordon, 1989, Theorem 1). Step (b) has used the fact that

$$\left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{R}_i^{t,\ell} - \boldsymbol{U}_i\right\|_{2,\infty} \leq \frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}},$$
$$\left\|\boldsymbol{X}_i^{t,\ell}\right\|_{2,\infty} \leq \left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{R}_i^{t,\ell} - \boldsymbol{U}_i\right\|_{2,\infty} + \|\boldsymbol{U}_i\|_{2,\infty} \leq \frac{9}{8}\sqrt{\frac{\mu r}{n}} \leq \frac{4}{3}\sqrt{\frac{\mu r}{n}}. \tag{121}$$

Similarly, one can obtain

$$\gamma_{1,b} \leq \frac{1}{2^{12}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}).$$

Combining the above bounds together gives that

$$\gamma_1 \leq \frac{1}{2^{11}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}). \tag{122}$$

C.3.2 Proof of (117)

Since $\mathcal{X}^{t,m}$ and $\boldsymbol{X}_i^{t,m}$ are independent of $\{\delta_{m,j}\}$ by construction, for any fixed $s \in [r^2]$,

$$\sum_{j=1}^{n^2}\left[\mathcal{M}_1(\mathcal{Z}^{t,m})\right]_{m,j}\left[\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right]_{j,s}$$
$$= \sum_{j\notin\Gamma_m}\left(p^{-1}\delta_{m,j} - 1\right)\left[\mathcal{M}_1(\mathcal{X}^{t,m} - \mathcal{T})\right]_{m,j}\left[\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right]_{j,s} := \sum_{j\notin\Gamma_m} X_j,$$

where Γ_m is an index set defined by

$$\Gamma_m = \{m, n+m, \cdots, n(m-2)+m, n(m-1)+1, \cdots, n(m-1)+n, nm+m, \cdots, n(n-1)+m\}.$$

Thus,

$$\gamma_2 = \sqrt{\sum_{s=1}^{r^2}\left(\sum_{j=1}^{n^2}\left[\mathcal{M}_1(\mathcal{Z}^{t,m})\right]_{m,j}\left[\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right]_{j,s}\right)^2}$$
$$\leq r \max_{1\leq s\leq r^2}\left|\sum_{j=1}^{n^2}\left[\mathcal{M}_1(\mathcal{Z}^{t,m})\right]_{m,j}\left[\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right]_{j,s}\right| = r \max_{1\leq s\leq r^2}\left|\sum_{j\notin\Gamma_m} X_j\right|.$$

A direct calculation gives that

$$|X_j| \leq p^{-1} \cdot \left\|\mathcal{X}^{t,m} - \mathcal{T}\right\|_\infty \cdot \left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^{t,m}\right\|_{2,\infty}\right)^2,$$

$$\left|\sum_{j\notin\Gamma_m}\mathbb{E}\left\{X_j^2\right\}\right| \le p^{-1}\sum_{j\notin\Gamma_m}\left|\left[\mathcal{M}_1(\mathcal{X}^{t,m}-\mathcal{T})\right]_{m,j}\right|^2\cdot\left|\left[\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right]_{j,s}\right|^2$$
$$\le p^{-1}\left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^{t,m}\right\|_{2,\infty}\right)^4\cdot n^2\left\|\mathcal{X}^{t,m}-\mathcal{T}\right\|_\infty^2.$$

Applying the matrix Bernstein inequality and taking a union bound over s shows that, with high probability,

$$\begin{aligned}\gamma_2 &\le C\cdot r\left(\frac{\log n}{p}+\sqrt{\frac{n^2\log n}{p}}\right)\left\|\mathcal{X}^{t,m}-\mathcal{T}\right\|_\infty\left(\max_{i=1,2,3}\left\|\boldsymbol{X}_i^{t,m}\right\|_{2,\infty}\right)^2\\ &\le C\cdot r\left(\frac{\log n}{p}+\sqrt{\frac{n^2\log n}{p}}\right)\cdot\frac{36}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\left(\frac{\mu r}{n}\right)^{3/2}\sigma_{\max}(\mathcal{T})\cdot 4\frac{\mu r}{n}\\ &\le \frac{1}{2^{11}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}),\end{aligned}\tag{123}$$

provided that $p\ge\frac{C_2\kappa^4\mu^4r^6\log n}{n^2}$.

C.3.3 Proof of (118)

Now we bound $\gamma_3=\left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{Z}^{t,m,\ell}\right)\left(\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right)\right\|_2$. To simplify notation, we denote by $\mathcal{H}_{\Omega_\ell}$ the operator $\mathcal{I}-p^{-1}\mathcal{P}_{\Omega_{-\ell}}-\mathcal{P}_\ell$ and define

$$\boldsymbol{C}_i^{t,m,\ell}=\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}-\boldsymbol{X}_i^{t,m}\boldsymbol{X}_i^{t,m\top}\quad\text{and}\quad\boldsymbol{D}_i^{t,m,\ell}=\boldsymbol{X}_i^{t,m}\boldsymbol{T}_i^{t,m}-\boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}.$$

Then $\boldsymbol{X}_i^{t,m}\boldsymbol{X}_i^{t,m\top}-\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}$ can be rewritten as

$$\boldsymbol{X}_i^{t,m}\boldsymbol{X}_i^{t,m\top}-\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}=\boldsymbol{D}_i^{t,m,\ell}\left(\boldsymbol{X}_i^{t,m}\boldsymbol{T}_i^{t,m}\right)^\top+\left(\boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right)\left(\boldsymbol{D}_i^{t,m,\ell}\right)^\top.\tag{124}$$

It is not hard to see that

$$\left\|\boldsymbol{D}_i^{t,m,\ell}\right\|_\mathsf{F}\le\left\|\boldsymbol{X}_i^{t,m}\boldsymbol{T}_i^{t,m}-\boldsymbol{X}_i^t\boldsymbol{R}_i^t\right\|_\mathsf{F}+\left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}-\boldsymbol{X}_i^t\boldsymbol{R}_i^t\right\|_\mathsf{F}\le 2\cdot\frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}},\tag{125}$$

where the last inequality is due to (17e). Recall that $\mathcal{Z}^{t,m,\ell}=\mathcal{H}_{\Omega_\ell}(\mathcal{X}^{t,\ell}-\mathcal{X}^{t,m})$. We have

$$\begin{aligned}\mathcal{X}^{t,\ell}-\mathcal{X}^{t,m}&=\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}-\left(\mathcal{T}+\mathcal{E}^{t-1,m}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{t,m}\boldsymbol{X}_i^{t,m\top}\\ &=\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}-\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{t,m}\boldsymbol{X}_i^{t,m\top}\\ &\quad+\left(\mathcal{E}^{t-1,\ell}-\mathcal{E}^{t-1,m}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{t,m}\boldsymbol{X}_i^{t,m\top}\\ &=\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\times_1\boldsymbol{C}_1^{t,m,\ell}\times_2\boldsymbol{X}_2^{t,m}\boldsymbol{X}_2^{t,m\top}\times_3\boldsymbol{X}_3^{t,m}\boldsymbol{X}_3^{t,m\top}\\ &\quad+\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\times_1\boldsymbol{X}_1^{t,m}\boldsymbol{X}_1^{t,m\top}\times_2\boldsymbol{C}_2^{t,m,\ell}\times_3\boldsymbol{X}_3^{t,m}\boldsymbol{X}_3^{t,m\top}\end{aligned}$$

$$
\begin{aligned}
&+\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\times_1 \boldsymbol{X}_1^{t,m}\boldsymbol{X}_1^{t,m\top}\times_2 \boldsymbol{X}_2^{t,m}\boldsymbol{X}_2^{t,m\top}\times_3 \boldsymbol{C}_3^{t,m,\ell}\\
&+\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\times_1 \boldsymbol{C}_1^{t,m,\ell}\times_2 \boldsymbol{C}_2^{t,m,\ell}\times_3 \boldsymbol{X}_3^{t,m}\boldsymbol{X}_3^{t,m\top}\\
&+\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\times_1 \boldsymbol{C}_1^{t,m,\ell}\times_2 \boldsymbol{X}_2^{t,m}\boldsymbol{X}_2^{t,m\top}\times_3 \boldsymbol{C}_3^{t,m,\ell}\\
&+\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\times_1 \boldsymbol{X}_1^{t,m}\boldsymbol{X}_1^{t,m\top}\times_2 \boldsymbol{C}_2^{t,m,\ell}\times_3 \boldsymbol{C}_3^{t,m,\ell}+\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{C}_i^{t,m,\ell}\\
&+\left(\mathcal{E}^{t-1,\ell}-\mathcal{E}^{t-1}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{t,m}\boldsymbol{X}_i^{t,m\top}+\left(\mathcal{E}^{t-1}-\mathcal{E}^{t-1,m}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{t,m}\boldsymbol{X}_i^{t,m\top}\\
:=&\sum_{i=1}^{9}\mathcal{Z}_i.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\gamma_3&=\left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{H}_{\Omega_\ell}\left(\mathcal{X}^{t,\ell}-\mathcal{X}^{t,m}\right)\right)\left(\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right)\right\|_2\\
&\le\sum_{i=1}^{9}\left\|\boldsymbol{e}_m^\top\mathcal{M}_1\left(\mathcal{H}_{\Omega_\ell}\left(\mathcal{Z}_i\right)\right)\left(\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right)\right\|_2:=\sum_{i=1}^{9}\gamma_{3,i}.
\end{aligned}
$$

Bounding $\gamma_{3,1}$. From (124), $\gamma_{3,1}$ can be decomposed as follows

$$
\begin{aligned}
\gamma_{3,1}\le&\left\|\boldsymbol{e}_m^\top\left(\mathcal{M}_1\left(\mathcal{H}_{\Omega_\ell}\left(\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\times_1\left(\boldsymbol{X}_1^{t,\ell}\boldsymbol{T}_1^{t,\ell}\right)\boldsymbol{D}_1^{t,m,\ell\top}\mathop{\times}_{i\neq1}\boldsymbol{X}_i^{t,m}\boldsymbol{X}_i^{t,m\top}\right)\right)\right)\left(\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right)\right\|_2\\
&+\left\|\boldsymbol{e}_m^\top\left(\mathcal{M}_1\left(\mathcal{H}_{\Omega_\ell}\left(\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\times_1\boldsymbol{D}_1^{t,m,\ell}\left(\boldsymbol{X}_1^{t,m}\boldsymbol{T}_1^{t,m}\right)^\top\mathop{\times}_{i\neq1}\boldsymbol{X}_i^{t,m}\boldsymbol{X}_i^{t,m\top}\right)\right)\right)\left(\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right)\right\|_2\\
=:&\ \gamma_{3,1}^a+\gamma_{3,1}^b.
\end{aligned}
$$

- Bounding $\gamma_{3,1}^a$. For simplification, we define

$$
\mathcal{C}_1:=\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\times_1\left(\boldsymbol{X}_1^{t,\ell}\boldsymbol{T}_1^{t,\ell}\right)\boldsymbol{D}_1^{t,m,\ell\top}\times_2\boldsymbol{X}_2^{t,m\top}\times_3\boldsymbol{X}_3^{t,m\top}.
$$

  It can be seen that

$$
\begin{aligned}
\left\|\mathcal{M}_1\left(\mathcal{C}_1\right)\right\|_{2,\infty}&\le\left\|\left(\boldsymbol{X}_1^{t,\ell}\boldsymbol{T}_1^{t,\ell}\right)\boldsymbol{D}_1^{t,m,\ell\top}\right\|_{2,\infty}\cdot\left\|\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\right\|\\
&\le\left\|\boldsymbol{X}_1^{t,\ell}\right\|_{2,\infty}\cdot\left\|\boldsymbol{D}_1^{t,m,\ell}\right\|_{\mathsf{F}}\cdot\left\|\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\right\|\\
&\le2\sqrt{\frac{\mu r}{n}}\cdot\frac{2}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\cdot2\sigma_{\max}(\mathcal{T}),
\end{aligned}
$$

  where the last line is due to $\left\|\boldsymbol{X}_i^{t,\ell}\right\|_{2,\infty}\le2\sqrt{\frac{\mu r}{n}}$, $\left\|\mathcal{M}_1\left(\mathcal{E}^{t,\ell}\right)\right\|\le\sigma_{\max}(\mathcal{T})$, and (125). Then the term $\gamma_{3,1}^a$ can be bounded as follows:

$$
\begin{aligned}
\gamma_{3,1}^a&\le\left\|\mathcal{M}_1\left(\mathcal{H}_{\Omega_\ell}\left(\mathcal{C}_1\times_2\boldsymbol{X}_2^{t,m}\times_3\boldsymbol{X}_3^{t,m}\right)\right)\left(\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right)\right\|_{\mathsf{F}}\\
&=\left\|\mathcal{H}_{\Omega_\ell}\left(\mathcal{C}_1\times_2\boldsymbol{X}_2^{t,m}\times_3\boldsymbol{X}_3^{t,m}\right)\times_2\boldsymbol{X}_2^{t,m\top}\times_3\boldsymbol{X}_3^{t,m\top}\right\|_{\mathsf{F}}
\end{aligned}
$$

$$
\begin{aligned}
&= \sup_{\mathcal{Z}\in\mathbb{R}^{n\times r\times r}:\|\mathcal{Z}\|_{\mathsf{F}}=1} \left\langle \mathcal{H}_{\Omega_\ell}\left(\mathcal{C}_1 \times_2 \boldsymbol{X}_2^{t,m} \times_3 \boldsymbol{X}_3^{t,m}\right) \times_2 \boldsymbol{X}_2^{t,m\mathsf{T}} \times_3 \boldsymbol{X}_3^{t,m\mathsf{T}}, \mathcal{Z}\right\rangle \\
&= \sup_{\mathcal{Z}\in\mathbb{R}^{n\times r\times r}:\|\mathcal{Z}\|_{\mathsf{F}}=1} \left\langle \mathcal{H}_{\Omega_\ell}\left(\mathcal{C}_1 \times_2 \boldsymbol{X}_2^{t,m} \times_3 \boldsymbol{X}_3^{t,m}\right), \mathcal{Z} \times_2 \boldsymbol{X}_2^{t,m} \times_3 \boldsymbol{X}_3^{t,m}\right\rangle \\
&\overset{(a)}{\leq} C\left(p^{-1}\log^3 n + \sqrt{p^{-1} n \log^5 n}\right) \cdot \left\|\mathcal{M}_1\left(\mathcal{C}_1\right)\right\|_{2,\infty} \prod_{i=2}^{3}\left(\left\|\boldsymbol{X}_i^{t,m}\right\|_{\mathsf{F}} \cdot \left\|\boldsymbol{X}_i^{t,m}\right\|_{2,\infty}\right) \\
&\leq C\left(p^{-1}\log^3 n + \sqrt{p^{-1} n \log^5 n}\right) \cdot \sqrt{\frac{\mu r}{n}} \frac{8}{2^{20}\kappa^2\mu^2 r^4} \frac{1}{2^t} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}) \cdot r \cdot 4\frac{\mu r}{n} \\
&\leq \frac{1}{2^{16}} \cdot \frac{1}{2^{20}\kappa^4\mu^2 r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}), \qquad (126)
\end{aligned}
$$

where step (a) follows from Lemma 30 and the last step has used the assumption

$$
p \geq \max\left\{\frac{C_1\kappa^2\mu^{1.5}r^{2.5}\log^3 n}{n^{3/2}}, \frac{C_2\kappa^4\mu^3 r^5\log^5 n}{n^2}\right\}
$$

- Bounding $\gamma_{3,1}^b$. Let $\boldsymbol{C}_2, \boldsymbol{C}_3$ be matrices defined by

$$
\begin{aligned}
&\boldsymbol{C}_2 := \boldsymbol{D}_1^{t,m,\ell}\left(\boldsymbol{X}_1^{t,m}\boldsymbol{T}_1^{t,m}\right)^{\mathsf{T}}\left(\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1,m}\right)\right)\left(\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right) \in \mathbb{R}^{n\times r^2}, \\
&\boldsymbol{C}_3 := \left(\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right)^{\mathsf{T}} \in \mathbb{R}^{r^2\times n^2}.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
\gamma_{3,1}^b &= \sqrt{\sum_{s=1}^{r^2}\left(\sum_{j=1, j\notin\Gamma}^{n^2}\sum_{k=1}^{r^2}\left[\boldsymbol{C}_2\right]_{m,k}\left[\boldsymbol{C}_3\right]_{k,j}\left(1-p^{-1}\delta_{m,j}\right)\left[\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right]_{j,s}\right)^2} \\
&= \sqrt{\sum_{s=1}^{r^2}\left(\sum_{k=1}^{r^2}\left[\boldsymbol{C}_2\right]_{m,k}\sum_{j=1, j\notin\Gamma}^{n^2}\left[\boldsymbol{C}_3\right]_{k,j}\left(1-p^{-1}\delta_{m,j}\right)\left[\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right]_{j,s}\right)^2} \\
&\leq \sqrt{\sum_{s=1}^{r^2}\sum_{k=1}^{r^2}\left[\boldsymbol{C}_2\right]_{m,k}^2 \cdot \sum_{k=1}^{r^2}\left(\sum_{j=1, j\notin\Gamma}^{n^2}\left[\boldsymbol{C}_3\right]_{k,j}\left(1-p^{-1}\delta_{m,j}\right)\left[\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right]_{j,s}\right)^2} \\
&\leq \left\|\boldsymbol{C}_2\right\|_{2,\infty} \cdot r^2 \cdot \max_{1\leq s,k\leq r^2}\left|\sum_{j=1, j\notin\Gamma}^{n^2}\left[\boldsymbol{C}_3\right]_{k,j}\left(1-p^{-1}\delta_{m,j}\right)\left[\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right]_{j,s}\right| \\
&:= \left\|\boldsymbol{C}_2\right\|_{2,\infty} \cdot r^2 \cdot \max_{1\leq s,k\leq r^2}\left|\sum_{j=1, j\notin\Gamma}^{n^2} X_j^{s,k}\right|.
\end{aligned}
$$

Since $\{\delta_{m,j}\}_{j\in[n^2]}$ and $\boldsymbol{X}_i^{t,m}$ are independent by construction, $X_j^{s,k}$ are independent mean zero random variables with

$$
\left|X_j^{s,k}\right| \leq p^{-1}\left\|\boldsymbol{C}_3\right\|_{\infty}\left\|\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right\|_{\infty} \leq p^{-1}\left\|\boldsymbol{X}_3^{t,m} \otimes \boldsymbol{X}_2^{t,m}\right\|_{2,\infty}^2,
$$

$$\left|\sum_{j=1,j\notin\Gamma}^{n^2}\mathbb{E}\left\{X_j^{s,k2}\right\}\right| \le p^{-1}\sum_{j=1}^{n^2}[\boldsymbol{C}_3]_{k,j}^2\left\|\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right\|_{2,\infty}^2 \le p^{-1}\left\|\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right\|_{2,\infty}^2.$$

We apply the Bernstein inequality to obtain that, with high probability,

$$\begin{aligned}\left|\sum_{j=1,j\notin\Gamma}^{n^2}X_j^{s,k}\right| &\le C\left(\frac{\log n}{p}\left\|\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right\|_{2,\infty}^2+\sqrt{\frac{\log n}{p}}\left\|\boldsymbol{X}_3^{t,m}\otimes\boldsymbol{X}_2^{t,m}\right\|_{2,\infty}\right)\\ &\le C\left(\frac{\log n}{p}\cdot 2^4\left(\frac{\mu r}{n}\right)^2+\sqrt{\frac{\log n}{p}}\cdot 2^2\frac{\mu r}{n}\right).\end{aligned}$$

Furthermore, a simple computation yields that

$$\begin{aligned}\|\boldsymbol{C}_2\|_{2,\infty} &\le \left\|\boldsymbol{D}_i^{t,m,\ell}\right\|_{2,\infty}\cdot\left\|\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1,m}\right)\right\|\\ &\le 2\max_{\ell\in[n]}\left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}-\boldsymbol{X}_i^t\boldsymbol{R}_i^t\right\|_{\mathsf{F}}\cdot\left\|\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1,m}\right)\right\|\\ &\le 4\sigma_{\max}(\mathcal{T})\cdot\max_{\ell\in[n]}\left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}-\boldsymbol{X}_i^t\boldsymbol{R}_i^t\right\|_{\mathsf{F}}.\end{aligned}$$

Thus we have

$$\begin{aligned}\gamma_{3,1}^b &\le \|\boldsymbol{C}_2\|_{2,\infty}\cdot r^2\cdot\max_{1\le s,k\le r^2}\left|\sum_{j=1,j\notin\Gamma}^{n^2}X_j^{s,k}\right|\\ &\le 4\sigma_{\max}(\mathcal{T})\max_{\ell\in[n]}\left\|\boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}-\boldsymbol{X}_i^t\boldsymbol{R}_i^t\right\|_{\mathsf{F}}\cdot r^2\cdot\max_{1\le s,k\le r^2}\left|\sum_{j=1,j\notin\Gamma}^{n^2}X_j^{s,k}\right|\\ &\le 4\sigma_{\max}(\mathcal{T})\frac{1}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\cdot r^2\cdot\left(\frac{\log n}{p}\cdot 2^4\left(\frac{\mu r}{n}\right)^2+\sqrt{\frac{\log n}{p}}\cdot 2^2\frac{\mu r}{n}\right)\\ &\le \frac{1}{2^{16}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \qquad (127)\end{aligned}$$

provided that $p\ge\frac{C_2\kappa^2\mu^2r^6\log n}{n^2}$.

Combining (126) and (127) together yields that

$$\gamma_{3,1}\le\frac{1}{2^{15}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}).$$

Bounding of $\gamma_{3,i}$ for $i=2,\cdots,9$. Following the same argument of bounding $\gamma_{3,1}^a$, one can obtain

$$\gamma_{3,i}\le\frac{1}{2^{15}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}),\quad i=2,\cdots,9.$$

Putting together all of the bounds on $\gamma_{3,i}$ for $i=1,\cdots,9$ yields that

$$\gamma_3\le\frac{1}{2^{11}}\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}).$$

C.4 Proof of Claim A.4

We only show the details for bounding $\left\|\boldsymbol{Y}_1^t\boldsymbol{Y}_1^t - \boldsymbol{Y}_1^{t,\ell}\boldsymbol{Y}_1^{t,\ell\top}\right\|_{\mathsf{F}}$, and the proofs for the other two cases are similar. Recall that $\mathcal{X}^t = \left(\mathcal{T}+\mathcal{E}^{t-1}\right) \times_{i=1}^{3} \boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top}$ and $\mathcal{X}^{t,\ell} = \left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right) \times_{i=1}^{3} \boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}$. In addition, $\boldsymbol{Y}_1^t$ and $\boldsymbol{Y}_1^{t,\ell}$ are the top-r eigenvectors of $\mathcal{M}_1^\top\left(\mathcal{X}^t\right)\mathcal{M}_1\left(\mathcal{X}^t\right)$ and $\mathcal{M}_1^\top\left(\mathcal{X}^{t,\ell}\right)\mathcal{M}_1\left(\mathcal{X}^{t,\ell}\right)$, respectively. Let $\boldsymbol{W}_1$, $\boldsymbol{D}_1$ and $\boldsymbol{D}_{3,2}$ be the matrices defined as

$$\boldsymbol{W}_1 := \mathcal{M}_1^\top\left(\mathcal{X}^t\right)\mathcal{M}_1\left(\mathcal{X}^t\right) - \mathcal{M}_1^\top\left(\mathcal{X}^{t,\ell}\right)\mathcal{M}_1\left(\mathcal{X}^{t,\ell}\right), \quad \boldsymbol{D}_1 := \boldsymbol{X}_1^t\boldsymbol{X}_1^{t\top} - \boldsymbol{X}_1^{t,\ell}\boldsymbol{X}_1^{t,\ell\top},$$
$$\boldsymbol{D}_{3,2} := \boldsymbol{X}_3^t\boldsymbol{X}_3^{t\top}\otimes\boldsymbol{X}_2^t\boldsymbol{X}_2^{t\top} - \boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\otimes\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}.$$

By the definition of $\mathcal{X}^t$ and $\mathcal{X}^{t,\ell}$, we have

$$\begin{aligned}
\|\boldsymbol{W}_1\|_{\mathsf{F}} \le& \left\|\boldsymbol{D}_{3,2}\mathcal{M}_1^\top\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\boldsymbol{X}_1^t\boldsymbol{X}_1^{t\top}\mathcal{M}_1\left(\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\times_{i\neq 1}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top}\right)\right\|_{\mathsf{F}} \\
&+\left\|\mathcal{M}_1^\top\left(\left(\mathcal{E}^{t-1}-\mathcal{E}^{t-1,\ell}\right)\times_{i\neq 1}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}\right)\boldsymbol{X}_1^t\boldsymbol{X}_1^{t\top}\mathcal{M}_1\left(\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\times_{i\neq 1}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top}\right)\right\|_{\mathsf{F}} \\
&+\left\|\mathcal{M}_1^\top\left(\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\times_{i\neq 1}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}\right)\boldsymbol{D}_1\mathcal{M}_1\left(\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\times_{i\neq 1}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top}\right)\right\|_{\mathsf{F}} \\
&+\left\|\mathcal{M}_1^\top\left(\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\times_{i\neq 1}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}\right)\boldsymbol{X}_1^{t,\ell}\boldsymbol{X}_1^{t,\ell\top}\mathcal{M}_1\left(\left(\mathcal{E}^{t-1}-\mathcal{E}^{t-1,\ell}\right)\times_{i\neq 1}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top}\right)\right\|_{\mathsf{F}} \\
&+\left\|\left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\otimes\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\right)\mathcal{M}_1^\top\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\boldsymbol{X}_1^{t,\ell}\boldsymbol{X}_1^{t,\ell\top}\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1,\ell}\right)\boldsymbol{D}_{3,2}\right\|_{\mathsf{F}} \\
\le& \|\boldsymbol{D}_{3,2}\|_{\mathsf{F}}\cdot\left(\left\|\boldsymbol{T}_1+\boldsymbol{E}_1^{t-1}\right\|^2+\left\|\boldsymbol{T}_1+\boldsymbol{E}_1^{t-1,\ell}\right\|^2\right)+\|\boldsymbol{D}_1\|_{\mathsf{F}}\cdot\left\|\boldsymbol{T}_1+\boldsymbol{E}_1^{t-1}\right\|\cdot\left\|\boldsymbol{T}_1+\boldsymbol{E}_1^{t-1,\ell}\right\| \\
&+\left\|\mathcal{E}^{t-1}-\mathcal{E}^{t-1,\ell}\right\|_{\mathsf{F}}\cdot\left(\left\|\boldsymbol{T}_1+\boldsymbol{E}_1^{t-1}\right\|+\left\|\boldsymbol{T}_1+\boldsymbol{E}_1^{t-1,\ell}\right\|\right) \\
\le& 8\sigma_{\max}(\mathcal{T})^2\|\boldsymbol{D}_{3,2}\|_{\mathsf{F}}+4\sigma_{\max}(\mathcal{T})\left\|\mathcal{E}^{t-1}-\mathcal{E}^{t-1,\ell}\right\|_{\mathsf{F}}+4\sigma_{\max}(\mathcal{T})^2\left\|\boldsymbol{X}_1^t\boldsymbol{X}_1^{t\top}-\boldsymbol{X}_1^{t,\ell}\boldsymbol{X}_1^{t,\ell\top}\right\|_{\mathsf{F}},
\end{aligned}$$

where the last line is due to (17a) and (17b). Furthermore, a straightforward computation yields that

$$\begin{aligned}
\|\boldsymbol{D}_{32}\|_{\mathsf{F}} \le& \left\|\boldsymbol{X}_3^t\boldsymbol{X}_3^{t\top}-\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\right\|_{\mathsf{F}}\cdot\left\|\boldsymbol{X}_2^t\boldsymbol{X}_2^{t\top}\right\|_{\mathsf{F}}+\left\|\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top}\right\|_{\mathsf{F}}\cdot\left\|\boldsymbol{X}_2^t\boldsymbol{X}_2^{t\top}-\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\right\|_{\mathsf{F}} \\
\le& 2r\cdot\max_{i=2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\top}-\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\top}\right\|_{\mathsf{F}}\le 4r\cdot\max_{i=2,3}\left\|\boldsymbol{X}_i^t\boldsymbol{R}_i^t-\boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right\|_{\mathsf{F}} \\
\le& \frac{1}{2^{18}\kappa^2\mu^2r^3}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}},
\end{aligned}$$

where the last inequality follows from (17e). Thus we can obtain

$$\begin{aligned}
\|\boldsymbol{W}_1\|_{\mathsf{F}} \le& \sigma_{\max}(\mathcal{T})^2\cdot\left(\frac{8}{2^{18}\kappa^2\mu^2r^3}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}+\frac{4}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}+\frac{8}{2^{20}\kappa^2\mu^2r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\right) \\
\le& \left(\frac{8}{2^{18}}+\frac{4}{2^{20}}+\frac{8}{2^{20}}\right)\frac{\sigma_{\max}^2(\mathcal{T})}{\kappa^2}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\le\frac{1}{2^{14}}\sigma_{\min}^2(\mathcal{T})\frac{1}{2^t}\sqrt{\frac{\mu r}{n}},
\end{aligned}$$

where the first inequality is due to (17d) and (17e). Consequently,

$$\|\boldsymbol{W}_1\| \leq \|\boldsymbol{W}_1\|_{\mathsf{F}} \leq \frac{1}{2^{14}}\sigma_{\min}^2(\mathcal{T}).$$

Note that the eigengap δ between the the r-th and $r+1$-th eigenvalues of $\mathcal{M}_1^{\mathsf{T}}(\mathcal{X}^{t,\ell})\mathcal{M}_1(\mathcal{X}^{t,\ell})$ is bounded by

$$\delta \geq \sigma_{\min}\left(\mathcal{M}_1\left(\mathcal{X}^{t,\ell}\right)\right)^2 \geq \left(\frac{15}{16}\right)^2 \sigma_{\min}^2(\mathcal{T}),$$

where the last inequality follows from Lemma 20. Applying Lemma 34 and Lemma 35 shows that

$$\left\|\boldsymbol{Y}_1^t\boldsymbol{Y}_1^{t\mathsf{T}} - \boldsymbol{Y}_1^{t,\ell}\boldsymbol{Y}_1^{t,\ell\mathsf{T}}\right\|_{\mathsf{F}} \leq \sqrt{2}\frac{\|\boldsymbol{W}_1\|_{\mathsf{F}}}{\delta - \|\boldsymbol{W}_1\|} \leq \frac{8\sqrt{2}}{7}\frac{1}{\sigma_{\min}^2(\mathcal{T})}\frac{1}{2^{14}}\sigma_{\min}^2(\mathcal{T})\frac{1}{2^t}\sqrt{\frac{\mu r}{n}} \leq \frac{1}{2^{13}}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}.$$

C.5 Proof of (91) in Claim A.5

We provide a detailed proof for $i=1$, while the proofs for $i=2,3$ are overall similar. Let $\boldsymbol{D}_i^{t,\ell}$ and $\boldsymbol{M}_i^{t,\ell}$ be two auxiliary matrices defined as

$$\begin{aligned}
\boldsymbol{D}_i^{t,\ell} &:= \boldsymbol{X}_i^t\boldsymbol{R}_i^t - \boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell},\\
\boldsymbol{M}_i^{t,\ell} &:= \boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}} - \boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\mathsf{T}} = \boldsymbol{D}_i^{t,\ell}\left(\boldsymbol{X}_i^t\boldsymbol{R}_i^t\right)^{\mathsf{T}} + \left(\boldsymbol{X}_i^{t,\ell}\boldsymbol{T}_i^{t,\ell}\right)\left(\boldsymbol{D}_i^{t,\ell}\right)^{\mathsf{T}}.
\end{aligned}$$

From (17c) and (17e), one has

$$\left\|\boldsymbol{D}_i^{t,\ell}\right\|_{\mathsf{F}} \leq \frac{1}{2^{20}\kappa^2\mu^2 r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}, \tag{128}$$

$$\left\|\boldsymbol{M}_i^{t,\ell}\right\|_{2,\infty} \leq \left\|\boldsymbol{M}_i^{t,\ell}\right\|_{\mathsf{F}} \leq 2\left\|\boldsymbol{D}_i^{t,\ell}\right\|_{\mathsf{F}} \leq \frac{1}{2^{19}\kappa^2\mu^2 r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}. \tag{129}$$

By the definition of $\mathcal{X}^t$ and $\mathcal{X}^{t,\ell}$,

$$\begin{aligned}
\mathcal{X}^t - \mathcal{X}^{t,\ell} &= \left(\mathcal{T} + \mathcal{E}^{t-1}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^t\boldsymbol{X}_i^{t\mathsf{T}} - \left(\mathcal{T} + \mathcal{E}^{t-1,\ell}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\mathsf{T}}\\
&= \left(\mathcal{T} + \mathcal{E}^{t-1}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{M}_i^{t,\ell} + \left(\mathcal{T} + \mathcal{E}^{t-1}\right)\mathop{\times}_{i\neq 3}\boldsymbol{M}_i^{t,\ell}\times_3\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\mathsf{T}}\\
&\quad + \left(\mathcal{T} + \mathcal{E}^{t-1}\right)\mathop{\times}_{i\neq 2}\boldsymbol{M}_i^{t,\ell}\times_2\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\mathsf{T}} + \left(\mathcal{T} + \mathcal{E}^{t-1}\right)\times_1\boldsymbol{X}_1^{t,\ell}\boldsymbol{X}_1^{t,\ell\mathsf{T}}\mathop{\times}_{i\neq 1}\boldsymbol{M}_i^{t,\ell}\\
&\quad + \left(\mathcal{T} + \mathcal{E}^{t-1}\right)\times_1\boldsymbol{M}_1^{t,\ell}\mathop{\times}_{i\neq 1}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\mathsf{T}} + \left(\mathcal{T} + \mathcal{E}^{t-1}\right)\mathop{\times}_{i\neq 2}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\mathsf{T}}\times_2\boldsymbol{M}_2^{t,\ell}\\
&\quad + \left(\mathcal{T} + \mathcal{E}^{t-1}\right)\mathop{\times}_{i\neq 3}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\mathsf{T}}\times_3\boldsymbol{M}_3^{t,\ell} + \left(\mathcal{E}^{t-1} - \mathcal{E}^{t-1,\ell}\right)\mathop{\times}_{i=1}^{3}\boldsymbol{X}_i^{t,\ell}\boldsymbol{X}_i^{t,\ell\mathsf{T}}\\
&:= \sum_{i=1}^{8}\mathcal{A}_i.
\end{aligned}$$

It follows that

$$\left\| \left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega \right) \left(\mathcal{X}^t - \mathcal{X}^{t,\ell} \right) \right) \underset{i \neq 1}{\times} \boldsymbol{X}_i^{t,\ell\top} \right\|_{\mathsf{F}} \le \sum_{i=1}^{8} \underbrace{\left\| \mathcal{M}_1 \left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega \right) (\mathcal{A}_i) \right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right) \right\|_{\mathsf{F}}}_{=:\eta_i}.$$

All of these terms but η_5 can be bounded by the same argument as controlling the term $\gamma_{3,1}^a$, yielding

$$\eta_i \le \frac{1}{2^{14}} \cdot \frac{1}{2^{20}\kappa^4\mu^2r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}), \quad i \neq 5. \tag{130}$$

For η_5, the same bound can be obtained, but a different strategy should be adopted.

C.5.1 Bounding η_5

Notice that

$$\boldsymbol{X}_i^{t,\ell} \boldsymbol{X}_i^{t,\ell\top} = \left(\boldsymbol{X}_i^{t,\ell} \boldsymbol{R}_i^{t,\ell} - \boldsymbol{U}_i \right) \left(\boldsymbol{X}_i^{t,\ell} \boldsymbol{R}_i^{t,\ell} \right)^\top + \boldsymbol{U}_i \left(\boldsymbol{X}_i^{t,\ell} \boldsymbol{R}_i^{t,\ell} - \boldsymbol{U}_i \right)^\top + \boldsymbol{U}_i \boldsymbol{U}_i^\top.$$

It follows that

$$\begin{aligned}
\mathcal{A}_5 &= \left(\mathcal{T} + \mathcal{E}^{t-1} \right) \times_1 \boldsymbol{M}_1^{t,\ell} \times_2 \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\top} \times_3 \boldsymbol{X}_3^{t,\ell} \boldsymbol{X}_3^{t,\ell\top} \\
&= \left(\mathcal{T} + \mathcal{E}^{t-1} \right) \times_1 \boldsymbol{D}_1^{t,\ell} \left(\boldsymbol{X}_1^{t} \boldsymbol{R}_1^{t} \right)^\top \times_2 \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\top} \times_3 \boldsymbol{X}_3^{t,\ell} \boldsymbol{X}_3^{t,\ell\top} \\
&\quad + \left(\mathcal{T} + \mathcal{E}^{t-1} \right) \times_1 \left(\boldsymbol{X}_1^{t,\ell} \boldsymbol{T}_1^{t,\ell} \right) \left(\boldsymbol{D}_1^{t,\ell} \right)^\top \times_2 \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\top} \times_3 \boldsymbol{X}_3^{t,\ell} \boldsymbol{X}_3^{t,\ell\top} \\
&= \left(\mathcal{T} + \mathcal{E}^{t-1} \right) \times_1 \boldsymbol{D}_1^{t,\ell} \left(\boldsymbol{X}_1^{t} \boldsymbol{R}_1^{t} \right)^\top \times_2 \left(\boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_2 \right) \left(\boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} \right)^\top \times_3 \boldsymbol{X}_3^{t,\ell} \boldsymbol{X}_3^{t,\ell\top} \\
&\quad + \left(\mathcal{T} + \mathcal{E}^{t-1} \right) \times_1 \boldsymbol{D}_1^{t,\ell} \left(\boldsymbol{X}_1^{t} \boldsymbol{R}_1^{t} \right)^\top \times_2 \boldsymbol{U}_2 \left(\boldsymbol{X}_2^{t,\ell} \boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_2 \right)^\top \times_3 \boldsymbol{X}_3^{t,\ell} \boldsymbol{X}_3^{t,\ell\top} \\
&\quad + \left(\mathcal{T} + \mathcal{E}^{t-1} \right) \times_1 \boldsymbol{D}_1^{t,\ell} \left(\boldsymbol{X}_1^{t} \boldsymbol{R}_1^{t} \right)^\top \times_2 \boldsymbol{U}_2 \boldsymbol{U}_2^\top \times_3 \left(\boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} - \boldsymbol{U}_3 \right) \left(\boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} \right)^\top \\
&\quad + \left(\mathcal{T} + \mathcal{E}^{t-1} \right) \times_1 \boldsymbol{D}_1^{t,\ell} \left(\boldsymbol{X}_1^{t} \boldsymbol{R}_1^{t} \right)^\top \times_2 \boldsymbol{U}_2 \boldsymbol{U}_2^\top \times_3 \boldsymbol{U}_3 \left(\boldsymbol{X}_3^{t,\ell} \boldsymbol{R}_3^{t,\ell} - \boldsymbol{U}_3 \right)^\top \\
&\quad + \left(\mathcal{T} + \mathcal{E}^{t-1} \right) \times_1 \boldsymbol{D}_1^{t,\ell} \left(\boldsymbol{X}_1^{t} \boldsymbol{R}_1^{t} \right)^\top \times_2 \boldsymbol{U}_2 \boldsymbol{U}_2^\top \times_3 \boldsymbol{U}_3 \boldsymbol{U}_3^\top \\
&\quad + \left(\mathcal{T} + \mathcal{E}^{t-1} \right) \times_1 \left(\boldsymbol{X}_1^{t,\ell} \boldsymbol{T}_1^{t,\ell} \right) \left(\boldsymbol{D}_1^{t,\ell} \right)^\top \times_2 \boldsymbol{X}_2^{t,\ell} \boldsymbol{X}_2^{t,\ell\top} \times_3 \boldsymbol{X}_3^{t,\ell} \boldsymbol{X}_3^{t,\ell\top} \\
&=: \sum_{i=1}^{6} \mathcal{B}_i,
\end{aligned}$$

which implies that η_5 can be expressed as

$$\eta_5 \le \sum_{i=1}^{6} \underbrace{\left\| \mathcal{M}_1 \left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega \right) (\mathcal{B}_i) \right) \left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell} \right) \right\|_{\mathsf{F}}}_{=:\eta_{5,i}}.$$

Bounding $\eta_{5,1},\ \eta_{5,2}, \eta_{5,3}$ and $\eta_{5,4}$. A simple computation yields that

$$\begin{aligned}
\mathcal{M}_1\left(\mathcal{B}_1\right) &= \mathcal{M}_1\left(\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\times_1 \boldsymbol{D}_1^{t,\ell}\left(\boldsymbol{X}_1^{t}\boldsymbol{R}_1^{t}\right)^{\mathsf{T}}\times_2\left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell}-\boldsymbol{U}_2\right)\left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell}\right)^{\mathsf{T}}\times_3 \boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\mathsf{T}}\right)\\
&= \boldsymbol{D}_1^{t,\ell}\left(\boldsymbol{X}_1^{t}\boldsymbol{R}_1^{t}\right)^{\mathsf{T}}\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\left(\left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\mathsf{T}}\right)\otimes\left(\left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell}-\boldsymbol{U}_2\right)\left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell}\right)^{\mathsf{T}}\right)\right)^{\mathsf{T}}\\
&= \underbrace{\boldsymbol{D}_1^{t,\ell}\left(\boldsymbol{X}_1^{t}\boldsymbol{R}_1^{t}\right)^{\mathsf{T}}\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell}\right)\right)}_{=:\boldsymbol{A}_1}\underbrace{\left(\boldsymbol{X}_3^{t,\ell}\otimes\left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell}-\boldsymbol{U}_2\right)\right)^{\mathsf{T}}}_{=:\boldsymbol{B}_1}.
\end{aligned}$$

Thus one has

$$\begin{aligned}
\eta_{5,1} &= \left\|\mathcal{M}_1\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{B}_1\right)\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)\right\|_{\mathsf{F}}\\
&= \left\|\left(\mathcal{M}_1\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right)\odot\mathcal{M}_1\left(\mathcal{B}_1\right)\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)\right\|_{\mathsf{F}}\\
&= \left\|\left(\mathcal{M}_1\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right)\odot\left(\boldsymbol{A}_1\boldsymbol{B}_1\right)\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)\right\|_{\mathsf{F}}\\
&= \sqrt{\sum_{i=1}^{n}\sum_{s=1}^{r^2}\left[\left(\mathcal{M}_1\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right)\odot\left(\boldsymbol{A}_1\boldsymbol{B}_1\right)\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)\right]_{is}^2}\\
&= \sqrt{\sum_{i=1}^{n}\sum_{s=1}^{r^2}\left(\sum_{j=1}^{n^2}\left[\mathcal{M}_1\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right)\odot\left(\boldsymbol{A}_1\boldsymbol{B}_1\right)\right]_{i,j}\cdot\left[\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right]_{j,s}\right)^2}\\
&= \sqrt{\sum_{i=1}^{n}\sum_{s=1}^{r^2}\left(\sum_{j=1}^{n^2}\left(1-p^{-1}\delta_{i,j}\right)\left(\sum_{q=1}^{r^2}\left[\boldsymbol{A}_1\right]_{i,q}\left[\boldsymbol{B}_1\right]_{q,j}\right)\cdot\left[\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right]_{j,s}\right)^2}\\
&= \sqrt{\sum_{i=1}^{n}\sum_{s=1}^{r^2}\left(\sum_{q=1}^{r^2}\left[\boldsymbol{A}_1\right]_{i,q}\sum_{j=1}^{n^2}\left(1-p^{-1}\delta_{i,j}\right)\left[\boldsymbol{B}_1\right]_{q,j}\cdot\left[\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right]_{j,s}\right)^2}\\
&\le \sqrt{\sum_{i=1}^{n}\sum_{s=1}^{r^2}\left(\sum_{q=1}^{r^2}\left[\boldsymbol{A}_1\right]_{i,q}^2\right)\cdot\left(\sum_{q=1}^{r^2}\left(\sum_{j=1}^{n^2}\left(1-p^{-1}\delta_{i,j}\right)\left[\boldsymbol{B}_1\right]_{q,j}\cdot\left[\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right]_{j,s}\right)^2\right)}\\
&\le \sqrt{\left(\sum_{i=1}^{n}\sum_{q=1}^{r^2}\left[\boldsymbol{A}_1\right]_{i,q}^2\right)\cdot\max_{i\in[n]}\sum_{s=1}^{r^2}\sum_{q=1}^{r^2}\left(\sum_{j=1}^{n^2}\left(1-p^{-1}\delta_{i,j}\right)\left[\boldsymbol{B}_1\right]_{q,j}\cdot\left[\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right]_{j,s}\right)^2}\\
&= \left\|\boldsymbol{A}_1\right\|_{\mathsf{F}}\cdot\sqrt{\max_{i\in[n]}\sum_{s=1}^{r^2}\sum_{q=1}^{r^2}\left(\sum_{j=1}^{n^2}\left(1-p^{-1}\delta_{i,j}\right)\left[\boldsymbol{B}_1\right]_{q,j}\cdot\left[\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right]_{j,s}\right)^2}\\
&\le \left\|\boldsymbol{A}_1\right\|_{\mathsf{F}}\cdot r^2\cdot\max_{i\in[n],s,q\in[r^2]}\left|\sum_{j=1}^{n^2}\left(1-p^{-1}\delta_{i,j}\right)\left[\boldsymbol{B}_1\right]_{q,j}\cdot\left[\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right]_{j,s}\right|
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \|\boldsymbol{A}_1\|_{\mathsf{F}} \cdot r^2 \cdot 3n^2 \cdot \|\boldsymbol{B}_1\|_{\infty} \cdot \left\|\boldsymbol{X}_2^{t,\ell}\right\|_{2,\infty} \left\|\boldsymbol{X}_3^{t,\ell}\right\|_{2,\infty} \\
&\leq \left\|\boldsymbol{D}_1^{t,\ell}\right\|_{\mathsf{F}} \cdot \left\|\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\right\| \cdot r^2 \cdot 3n^2 \cdot \left\|\boldsymbol{X}_3^{t,\ell}\right\|_{2,\infty}^2 \cdot \left\|\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_2\right\|_{2,\infty} \cdot \left\|\boldsymbol{X}_2^{t,\ell}\right\|_{2,\infty} \\
&\overset{(b)}{\leq} \frac{1}{2^{20}\kappa^2\mu^2 r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}} \cdot \frac{9}{8}\sigma_{\max}(\mathcal{T}) \cdot 3n^2 r^2 \cdot \left(\frac{9}{8}\right)^3 \left(\frac{\mu r}{n}\right)^{3/2} \cdot \frac{1}{2^{20}\kappa^2\mu^2 r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}} \\
&\leq \frac{1}{6}\frac{1}{2^{14}} \cdot \frac{1}{2^{20}\kappa^4\mu^2 r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}),
\end{aligned}
\tag{131}
$$

where $\odot$ denotes the entrywise product, step (a) uses the same argument as in (120) and step (b) follows from the inequalities (121) and

$$
\left\|\mathcal{M}_1(\mathcal{T}+\mathcal{E}^{t-1})\right\| \leq \sigma_{\max}(\mathcal{T}) + \frac{1}{2^{20}\kappa^6\mu^2 r^4}\frac{1}{2^t}\sigma_{\max}(\mathcal{T}) \leq \frac{9}{8}\sigma_{\max}(\mathcal{T}).
$$

Applying the same argument as above can show that

$$
\eta_{5,2} \leq \frac{1}{6}\frac{1}{2^{14}} \cdot \frac{1}{2^{20}\kappa^4\mu^2 r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \tag{132}
$$

$$
\eta_{5,3} \leq \frac{1}{6}\frac{1}{2^{14}} \cdot \frac{1}{2^{20}\kappa^4\mu^2 r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \tag{133}
$$

$$
\eta_{5,4} \leq \frac{1}{6}\frac{1}{2^{14}} \cdot \frac{1}{2^{20}\kappa^4\mu^2 r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}). \tag{134}
$$

Bounding $\eta_{5,5}$. Notice that

$$
\begin{aligned}
\mathcal{M}_1\left(\mathcal{B}_5\right) &= \mathcal{M}_1\left(\left(\mathcal{T}+\mathcal{E}^{t-1}\right) \times_1 \boldsymbol{D}_1^{t,\ell}\left(\boldsymbol{X}_1^t\boldsymbol{R}_1^t\right)^{\top} \times_2 \boldsymbol{U}_2\boldsymbol{U}_2^{\top} \times_3 \boldsymbol{U}_3\boldsymbol{U}_3^{\top}\right) \\
&= \underbrace{\boldsymbol{D}_1^{t,\ell}\left(\boldsymbol{X}_1^t\boldsymbol{R}_1^t\right)^{\top}\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\left(\boldsymbol{U}_3\otimes\boldsymbol{U}_2\right)}_{=:\boldsymbol{A}_5}\underbrace{\left(\boldsymbol{U}_3\otimes\boldsymbol{U}_2\right)^{\top}}_{=:\boldsymbol{B}_5},
\end{aligned}
$$

where the matrices $\boldsymbol{A}_5$ and $\boldsymbol{B}_5$ obey that

$$
\begin{aligned}
\|\boldsymbol{A}_5\|_{\mathsf{F}} &\leq \left\|\boldsymbol{D}_1^{t,\ell}\right\|_{\mathsf{F}} \cdot \left\|\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\right\| \leq \frac{9}{2^{23}\kappa^2\mu^2 r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}), \\
\|\boldsymbol{B}_5\|_{\infty} &\leq \|\boldsymbol{U}_2\|_{2,\infty} \cdot \|\boldsymbol{U}_3\|_{2,\infty} \leq \frac{\mu r}{n}.
\end{aligned}
$$

A simple computation yields that

$$
\begin{aligned}
\eta_{5,5} &= \left\|\mathcal{M}_1\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega}\right)(\mathcal{B}_5)\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)\left(\boldsymbol{R}_3^{t,\ell}\otimes\boldsymbol{R}_2^{t,\ell}\right)\right\|_{\mathsf{F}} \\
&\leq \left\|\mathcal{M}_1\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega}\right)(\mathcal{B}_5)\right)\left(\left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{R}_3^{t,\ell}\right)\otimes\left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell}\right) - \boldsymbol{U}_3\otimes\boldsymbol{U}_2\right)\right\|_{\mathsf{F}} \\
&\quad + \left\|\mathcal{M}_1\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega}\right)(\mathcal{B}_5)\right)\left(\boldsymbol{U}_3\otimes\boldsymbol{U}_2\right)\right\|_{\mathsf{F}} \\
&\leq \underbrace{\left\|\mathcal{M}_1\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega}\right)(\mathcal{B}_5)\right)\left(\left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{R}_3^{t,\ell}-\boldsymbol{U}_3\right)\otimes\left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell}\right)\right)\right\|_{\mathsf{F}}}_{=:\eta_{5,5}^a}
\end{aligned}
$$

$$+ \underbrace{\left\| \mathcal{M}_1\left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega \right)(\mathcal{B}_5) \right) \left(\boldsymbol{U}_3 \otimes \left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_2 \right) \right) \right\|_{\mathsf{F}}}_{=:\eta_{5,5}^b}$$

$$+ \underbrace{\left\| \mathcal{M}_1\left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega \right)(\mathcal{B}_5) \right) (\boldsymbol{U}_3 \otimes \boldsymbol{U}_2) \right\|_{\mathsf{F}}}_{=:\eta_{5,5}^c}.$$

- Controlling $\eta_{5,5}^a$. It can be bounded as follows:

$$\begin{aligned}
\eta_{5,5}^a &= \left\| \mathcal{M}_1\left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega \right)(\mathcal{B}_5) \right) \left(\left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{R}_3^{t,\ell} - \boldsymbol{U}_3 \right) \otimes \left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell} \right) \right) \right\|_{\mathsf{F}} \\
&\leq \|\boldsymbol{A}_5\|_{\mathsf{F}} \cdot r^2 \cdot \max_{i\in[n],s,q\in[r^2]} \left| \sum_{j=1}^{n^2} \left(1 - p^{-1}\delta_{i,j} \right) [\boldsymbol{B}_5]_{q,j} \cdot \left[\left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{R}_3^{t,\ell} - \boldsymbol{U}_3 \right) \otimes \left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell} \right) \right]_{j,s} \right| \\
&\overset{(a)}{\leq} \|\boldsymbol{A}_5\|_{\mathsf{F}} \cdot r^2 \cdot 3n^2 \|\boldsymbol{B}_5\|_\infty \cdot \left\| \left(\boldsymbol{X}_3^{t,\ell}\boldsymbol{R}_3^{t,\ell} - \boldsymbol{U}_3 \right) \otimes \left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell} \right) \right\|_\infty \\
&\leq \|\boldsymbol{A}_5\|_{\mathsf{F}} \cdot r^2 \cdot 3n^2 \|\boldsymbol{B}_5\|_\infty \cdot \left\| \boldsymbol{X}_3^{t,\ell}\boldsymbol{R}_3^{t,\ell} - \boldsymbol{U}_3 \right\|_{2,\infty} \cdot \left\| \boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell} \right\|_{2,\infty} \\
&\leq \frac{9}{2^{23}\kappa^2\mu^2r^4} \frac{1}{2^t} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}) \cdot 3n^2r^2 \cdot \frac{\mu r}{n} \cdot \frac{1}{2^{20}\kappa^2\mu^2r^4} \frac{1}{2^t} \sqrt{\frac{\mu r}{n}} \cdot \frac{9}{8} \sqrt{\frac{\mu r}{n}} \\
&\leq \frac{1}{18}\frac{1}{2^{14}} \cdot \frac{1}{2^{20}\kappa^4\mu^2r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}),
\end{aligned}$$

where step (a) uses the same argument as in (120).

- Controlling $\eta_{5,5}^b$. Similarly, one has

$$\begin{aligned}
\eta_{5,5}^b &\leq \|\boldsymbol{A}_5\|_{\mathsf{F}} \cdot r^2 \cdot \max_{i\in[n],s,q\in[r^2]} \left| \sum_{j=1}^{n^2} \left(1 - p^{-1}\delta_{i,j} \right) [\boldsymbol{B}_5]_{q,j} \cdot \left[\boldsymbol{U}_3 \otimes \left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_2 \right) \right]_{j,s} \right| \\
&\leq \|\boldsymbol{A}_5\|_{\mathsf{F}} \cdot r^2 \cdot 3n^2 \|\boldsymbol{B}_5\|_\infty \cdot \left\| \boldsymbol{U}_3 \otimes \left(\boldsymbol{X}_2^{t,\ell}\boldsymbol{R}_2^{t,\ell} - \boldsymbol{U}_2 \right) \right\|_\infty \\
&\leq \frac{1}{18}\frac{1}{2^{14}} \cdot \frac{1}{2^{20}\kappa^4\mu^2r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}).
\end{aligned}$$

- Controlling $\eta_{5,5}^c$. Recall that

$$\begin{aligned}
\mathcal{B}_5 &= \left(\mathcal{T} + \mathcal{E}^{t-1} \right) \times_1 \boldsymbol{D}_1^{t,\ell} \left(\boldsymbol{X}_1^t\boldsymbol{R}_1^t \right)^{\mathsf{T}} \times_2 \boldsymbol{U}_2\boldsymbol{U}_2{}^{\mathsf{T}} \times_3 \boldsymbol{U}_3\boldsymbol{U}_3{}^{\mathsf{T}} \\
&= \underbrace{\left(\left(\mathcal{T} + \mathcal{E}^{t-1} \right) \times_1 \boldsymbol{D}_1^{t,\ell} \left(\boldsymbol{X}_1^t\boldsymbol{R}_1^t \right)^{\mathsf{T}} \times_2 \boldsymbol{U}_2^{\mathsf{T}} \times_3 \boldsymbol{U}_3^{\mathsf{T}} \right)}_{=:\mathcal{Y}_{5,5}} \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{U}_3,
\end{aligned}$$

where $\mathcal{Y}_{5,5} \in \mathbb{R}^{n\times r\times r}$. A straightforward computation yields that

$$\begin{aligned}
\eta_{5,5}^c &= \left\| \mathcal{M}_1\left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega \right)(\mathcal{B}_5) \right) (\boldsymbol{U}_3 \otimes \boldsymbol{U}_2) \right\|_{\mathsf{F}} \\
&= \left\| \left(\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega \right)(\mathcal{B}_5) \right) \times_2 \boldsymbol{U}_2^{\mathsf{T}} \times_3 \boldsymbol{U}_3^{\mathsf{T}} \right\|_{\mathsf{F}}
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{\mathcal{Z}\in\mathbb{R}^{n\times n\times n}:\|\mathcal{Z}\|_{\mathsf{F}}=1} \left\langle \left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega}\right)\left(\mathcal{Y}_{5,5}\times_2 \boldsymbol{U}_2\times_3 \boldsymbol{U}_3\right), \mathcal{Z}\times_2 \boldsymbol{U}_2\times_3 \boldsymbol{U}_3\right\rangle \\
&\overset{(a)}{\leq} C\sqrt{\frac{\mu^2 r^2\log n}{n^2 p}}\cdot\left\|\mathcal{Y}_{5,5}\right\|_{\mathsf{F}} = C\sqrt{\frac{\mu^2 r^2\log n}{n^2 p}}\cdot\left\|\mathcal{M}_1\left(\mathcal{Y}_{5,5}\right)\right\|_{\mathsf{F}} \\
&\leq C\sqrt{\frac{\mu^2 r^2\log n}{n^2 p}}\cdot\left\|\boldsymbol{D}_1^{t,\ell}\right\|_{\mathsf{F}}\cdot\left\|\boldsymbol{X}_1^{t}\right\|\cdot\left\|\mathcal{M}_1\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\right\| \\
&\leq C\sqrt{\frac{\mu^2 r^2\log n}{n^2 p}}\cdot\frac{1}{2^{20}\kappa^2\mu^2 r^4}\frac{1}{2^t}\sqrt{\frac{\mu r}{n}}\cdot\frac{9}{8}\sigma_{\max}\left(\mathcal{T}\right) \\
&\leq \frac{1}{18}\frac{1}{2^{14}}\cdot\frac{1}{2^{20}\kappa^4\mu^2 r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}),
\end{aligned}
$$

provided that $p \geq \frac{C_2\kappa^4\mu^2 r^2\log n}{n^2}$, where (a) follows from Lemma 27 .

Combining the above bounds together, we have

$$
\eta_{5,5} \leq \frac{1}{6}\frac{1}{2^{14}}\cdot\frac{1}{2^{20}\kappa^4\mu^2 r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}). \tag{135}
$$

Bounding $\eta_{5,6}$. Notice that

$$
\begin{aligned}
\mathcal{B}_6 &= \left(\mathcal{T}+\mathcal{E}^{t-1}\right)\times_1\left(\boldsymbol{X}_1^{t,\ell}\boldsymbol{T}_1^{t,\ell}\right)\left(\boldsymbol{D}_1^{t,\ell}\right)^{\top}\times_2\boldsymbol{X}_2^{t,\ell}\boldsymbol{X}_2^{t,\ell\top}\times_3\boldsymbol{X}_3^{t,\ell}\boldsymbol{X}_3^{t,\ell\top} \\
&= \underbrace{\left(\left(\mathcal{T}+\mathcal{E}^{t-1}\right)\times_1\left(\boldsymbol{X}_1^{t,\ell}\boldsymbol{T}_1^{t,\ell}\right)\left(\boldsymbol{D}_1^{t,\ell}\right)^{\top}\times_2\boldsymbol{X}_2^{t,\ell\top}\times_3\boldsymbol{X}_3^{t,\ell\top}\right)}_{=:\mathcal{Y}_{5,6}}\times_2\boldsymbol{X}_2^{t,\ell}\times_3\boldsymbol{X}_3^{t,\ell},
\end{aligned}
$$

where $\mathcal{Y}_{5,6}\in\mathbb{R}^{n\times r\times r}$. Following the same argument of bounding $\gamma_{3,1}^{a}$, one has

$$
\eta_{5,6} \leq \frac{1}{6}\frac{1}{2^{14}}\cdot\frac{1}{2^{20}\kappa^4\mu^2 r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}). \tag{136}
$$

Combining $\eta_{5,i}$ together. Combining (131), (132), (133), (134), (135) and (136) together, we have

$$
\eta_5 \leq \sum_{i=1}^{6}\eta_{5,i} \leq \frac{1}{2^{14}}\cdot\frac{1}{2^{20}\kappa^4\mu^2 r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}). \tag{137}
$$

C.5.2 Combining η_i together

Combining (130) and (137) together yields that

$$
\left\|\left(\left(\mathcal{I}-p^{-1}\mathcal{P}_{\Omega}\right)\left(\mathcal{X}^{t}-\mathcal{X}^{t,\ell}\right)\right)\underset{j\neq 1}{\times}\boldsymbol{X}_j^{t,\ell\top}\right\|_{\mathsf{F}} \leq \sum_{i=1}^{8}\eta_i \leq \frac{1}{2^{11}}\cdot\frac{1}{2^{20}\kappa^4\mu^2 r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}).
$$

Thus we have completed the proof of (91) in Claim A.5.

C.6 Proof of (92) in Claim A.5

We prove the case $i = 1$ and the others are similar. Notice that only the ℓ-th row and those columns in the set Γ of

$$\mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right)$$

are non-zeros. By the definition of $\mathcal{P}_{\ell,:}$ and $\mathcal{P}_{-\ell,\Gamma}$ in (36) and (37), we have

$$\begin{aligned}
&\left\|\left(\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right) \underset{j\neq 1}{\times} \boldsymbol{X}_j^{t,\ell\mathsf{T}}\right\|_{\mathsf{F}} \\
&= \left\|\mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right)\left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell}\right)\right\|_{\mathsf{F}} \\
&\leq \left\|\boldsymbol{e}_\ell^{\mathsf{T}}\mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right)\left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell}\right)\right\|_2 \\
&\quad + \left\|\mathcal{P}_{-\ell,\Gamma}\left(\mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right)\right)\left(\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell}\right)\right\|_{\mathsf{F}} := \phi_1 + \phi_2.
\end{aligned}$$

C.6.1 Bounding ϕ_1

A straightforward computation yields that

$$\begin{aligned}
\phi_1 &= \left\|\sum_{j=1}^{n^2}\left[\mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right)\right]_{\ell,j}\left[\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell}\right]_{j,:}\right\|_2 \\
&= \left\|\sum_{j=1}^{n^2}\left(1 - p^{-1}\delta_{\ell,j}\right)\left[\mathcal{M}_1\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right]_{\ell,j}\left[\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell}\right]_{j,:}\right\|_2 := \left\|\sum_{j=1}^{n^2}\boldsymbol{x}_j^{\mathsf{T}}\right\|_2.
\end{aligned}$$

Notice that conditioned on $\boldsymbol{X}_i^{t,\ell}$ for $i = 2, 3$, the vectors $\boldsymbol{x}_j \in \mathbb{R}^{r^2\times 1}$ are independent mean-zero random vectors with

$$\begin{aligned}
\|\boldsymbol{x}_j\|_2 &\leq \frac{1}{p}\left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\infty \cdot \left\|\boldsymbol{X}_2^{t,\ell}\right\|_{2,\infty} \cdot \left\|\boldsymbol{X}_3^{t,\ell}\right\|_{2,\infty}, \\
\left\|\sum_{j=1}^{n^2}\mathbb{E}\left\{\boldsymbol{x}_j\boldsymbol{x}_j^{\mathsf{T}}\right\}\right\| &\leq p^{-1}\sum_{j=1}^{n^2}\left[\mathcal{M}_1\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right]_{\ell,j}^2 \cdot \left\|\left[\boldsymbol{X}_3^{t,\ell} \otimes \boldsymbol{X}_2^{t,\ell}\right]_{j,:}\right\|_2^2 \\
&\leq \frac{n^2}{p} \cdot \left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\infty^2 \cdot \left\|\boldsymbol{X}_2^{t,\ell}\right\|_{2,\infty}^2 \cdot \left\|\boldsymbol{X}_3^{t,\ell}\right\|_{2,\infty}^2, \\
\left|\sum_{j=1}^{n^2}\mathbb{E}\left\{\boldsymbol{x}_j^{\mathsf{T}}\boldsymbol{x}_j\right\}\right| &\leq \frac{n^2}{p} \cdot \left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\infty^2 \cdot \left\|\boldsymbol{X}_2^{t,\ell}\right\|_{2,\infty}^2 \cdot \left\|\boldsymbol{X}_3^{t,\ell}\right\|_{2,\infty}^2.
\end{aligned}$$

By the matrix Bernstein inequality, with high probability, one has

$$\phi_1 \leq C\left(\frac{\log n}{p} + \sqrt{\frac{n^2\log n}{p}}\right)\left\|\mathcal{X}^{t,\ell} - \mathcal{T}\right\|_\infty \cdot \left\|\boldsymbol{X}_2^{t,\ell}\right\|_{2,\infty} \cdot \left\|\boldsymbol{X}_3^{t,\ell}\right\|_{2,\infty}$$

$$\leq C\left(\frac{\log n}{p}+\sqrt{\frac{n^2\log n}{p}}\right)\cdot\frac{36}{2^{20}\kappa^2\mu^2r^4}\cdot\frac{1}{2^t}\cdot\left(\frac{\mu r}{n}\right)^{3/2}\sigma_{\max}(\mathcal{T})\cdot\left(\frac{9}{8}\right)^2\frac{\mu r}{n}$$
$$\leq\frac{1}{2^{12}}\cdot\frac{1}{2^{20}\kappa^4\mu^2r^4}\frac{1}{2^{t+1}}\sqrt{\frac{\mu r}{n}}\sigma_{\max}(\mathcal{T}) \tag{138}$$

under the assumption $p\geq\frac{C_2\kappa^4\mu^4r^4\log n}{n^2}$.

C.6.2 Bounding ϕ_2

A simple computation yields that

$$\begin{aligned}\phi_2&=\left\|\mathcal{P}_{-\ell,\Gamma}\left(\mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}}+\mathcal{P}_\ell-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^{t,\ell}-\mathcal{T}\right)\right)\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)\right\|_{\mathsf{F}}\\&\leq r\left\|\mathcal{P}_{-\ell,\Gamma}\left(\mathcal{M}_1\left(\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}}+\mathcal{P}_\ell-p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^{t,\ell}-\mathcal{T}\right)\right)\right)\left(\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right)\right\|\\&=r\left\|\sum_{j\in\Gamma}\boldsymbol{a}_j\boldsymbol{b}_j^\top\right\|:=r\left\|\sum_{j\in\Gamma}\boldsymbol{Z}_j\right\|,\end{aligned}$$

where $\boldsymbol{a}_j\in\mathbb{R}^{n\times1}$ with $[\boldsymbol{a}_j]_k=\left(1-p^{-1}\delta_{k,j}\right)\left[\mathcal{M}_1\left(\mathcal{X}^{t,\ell}-\mathcal{T}\right)\right]_{k,j}$ for $k\neq\ell$ and $[\boldsymbol{a}_j]_\ell=0$, and $\boldsymbol{b}_j=\left[\boldsymbol{X}_3^{t,\ell}\otimes\boldsymbol{X}_2^{t,\ell}\right]_{j,:}^\top\in\mathbb{R}^{r^2\times1}$. Notice that conditioned on $\boldsymbol{X}_i^{t,\ell}$, $\boldsymbol{Z}_j$ are independent mean-zero matrices with

$$\|\boldsymbol{Z}_j\|\leq\|\boldsymbol{a}_j\|_2\cdot\|\boldsymbol{b}_j\|_2\leq\frac{\sqrt{n}}{p}\cdot\left\|\mathcal{X}^{t,\ell}-\mathcal{T}\right\|_\infty\cdot\left\|\boldsymbol{X}_2^{t,\ell}\right\|_{2,\infty}\cdot\left\|\boldsymbol{X}_3^{t,\ell}\right\|_{2,\infty},$$
$$\begin{aligned}\left\|\sum_{j\in\Gamma}\mathbb{E}\left\{\boldsymbol{Z}_j\boldsymbol{Z}_j^\top\right\}\right\|&=\left\|\sum_{j\in\Gamma}\mathbb{E}\left\{\boldsymbol{a}_j\boldsymbol{a}_j^\top\right\}\|\boldsymbol{b}_j\|_2^2\right\|\leq\max_{i\in[n]}\sum_{j\in\Gamma}p^{-1}\left[\mathcal{M}_1\left(\mathcal{X}^{t,\ell}-\mathcal{T}\right)\right]_{i,j}^2\cdot\|\boldsymbol{b}_j\|_2^2\\&\leq\frac{2n}{p}\cdot\left\|\mathcal{X}^{t,\ell}-\mathcal{T}\right\|_\infty^2\cdot\left\|\boldsymbol{X}_2^{t,\ell}\right\|_{2,\infty}^2\cdot\left\|\boldsymbol{X}_3^{t,\ell}\right\|_{2,\infty}^2,\end{aligned}$$

where we have used the fact that $|\Gamma|\leq 2n$. Similarly, one has

$$\begin{aligned}\left\|\sum_{j\in\Gamma}\mathbb{E}\left\{\boldsymbol{Z}_j^\top\boldsymbol{Z}_j\right\}\right\|&=\left\|\sum_{j\in\Gamma}\mathbb{E}\left\{\|\boldsymbol{a}_j\|_2^2\right\}\boldsymbol{b}_j\boldsymbol{b}_j^\top\right\|\leq\sum_{j\in\Gamma}\|\boldsymbol{b}_j\|_2^2\cdot\mathbb{E}\left\{\|\boldsymbol{a}_j\|_2^2\right\}\\&\leq\sum_{j\in\Gamma}\|\boldsymbol{b}_j\|_2^2\cdot p^{-1}\sum_{i=1}^n\left[\mathcal{M}_1\left(\mathcal{X}^{t,\ell}-\mathcal{T}\right)\right]_{i,j}^2\\&\leq\frac{2n^2}{p}\left\|\mathcal{X}^{t,\ell}-\mathcal{T}\right\|_\infty^2\cdot\left\|\boldsymbol{X}_2^{t,\ell}\right\|_{2,\infty}^2\cdot\left\|\boldsymbol{X}_3^{t,\ell}\right\|_{2,\infty}^2.\end{aligned}$$

Applying the matrix Bernstein inequality shows that, with high probability,

$$\phi_2\leq r\cdot C\left(\frac{\sqrt{n}\log n}{p}+\sqrt{\frac{2n^2\log n}{p}}\right)\cdot\left\|\mathcal{X}^{t,\ell}-\mathcal{T}\right\|_\infty\cdot\left\|\boldsymbol{X}_2^{t,\ell}\right\|_{2,\infty}\cdot\left\|\boldsymbol{X}_3^{t,\ell}\right\|_{2,\infty}$$

$$\begin{aligned}
&\le r \cdot C\left(\frac{\sqrt{n}\log n}{p} + \sqrt{\frac{n^2\log n}{p}}\right) \cdot \frac{36}{2^{20}\kappa^2\mu^2r^4} \cdot \frac{1}{2^t} \cdot \left(\frac{\mu r}{n}\right)^{3/2} \sigma_{\max}(\mathcal{T}) \cdot \left(\frac{9}{8}\right)^2 \frac{\mu r}{n} \\
&\le \frac{1}{2^{12}} \cdot \frac{1}{2^{20}\kappa^4\mu^2r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}), \qquad (139)
\end{aligned}$$

provided that $p \ge \max\left\{\frac{C_1\kappa^2\mu^2r^3\log n}{n^{1.5}}, \frac{C_2\kappa^4\mu^4r^6\log n}{n^2}\right\}$.

Combining (138) and (139) together gives

$$\left\| \left(\left(p^{-1}\mathcal{P}_{\Omega_{-\ell}} + \mathcal{P}_\ell - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{X}^{t,\ell} - \mathcal{T}\right)\right) \underset{j\neq 1}{\times} \boldsymbol{X}_j^{t,\ell\mathsf{T}} \right\|_{\mathsf{F}} \le \frac{1}{2^{11}} \cdot \frac{1}{2^{20}\kappa^4\mu^2r^4} \frac{1}{2^{t+1}} \sqrt{\frac{\mu r}{n}} \sigma_{\max}(\mathcal{T}),$$

which completes the proof of (92) in Claim A.5.

Appendix D. Auxiliary Lemmas

Lemma 21 (Xia and Yuan, 2019, Lemma 1) *For any $\mathcal{X} \in \mathbb{R}^{n\times n\times n}$ with multilinear rank (r_1, r_2, r_3), one has*

$$\|\mathcal{X}\|_\infty \le \|\mathcal{X}\| \le \|\mathcal{X}\|_{\mathsf{F}} \le \sqrt{r_1r_2r_3} \cdot \|\mathcal{X}\| \quad \text{and} \quad \|\mathcal{X}\|_* \le \min\left\{\sqrt{r_1r_2}, \sqrt{r_1r_3}, \sqrt{r_2r_3}\right\} \cdot \|\mathcal{X}\|_{\mathsf{F}}.$$

Lemma 22 (Xia et al., 2017, Lemma 6) *For a tensor $\mathcal{X} \in \mathbb{R}^{n\times n\times n}$ with multilinear rank $\boldsymbol{r} = (r_1, r_2, r_3)$, the following bound holds for $i = 1, 2, 3$*

$$\|\mathcal{M}_i(\mathcal{X})\| \le \sqrt{\frac{\prod_{j=1}^3 r_j}{r_i \cdot \max_{j'\neq i} r_{j'}}} \cdot \|\mathcal{X}\| = \sqrt{\min_{j\neq i} r_j} \cdot \|\mathcal{X}\|.$$

Lemma 23 (Cai et al., 2021b, Lemma EC.12; Tong et al., 2021, Lemma 13) *Suppose Ω satisfies the Bernoulli observation model. Then for any fixed $\mathcal{Z} \in \mathbb{R}^{n\times n\times n}$,*

$$\left\|\left(p^{-1}\mathcal{P}_\Omega - \mathcal{I}\right)(\mathcal{Z})\right\| \le C\left(p^{-1}\log^3 n \|\mathcal{Z}\|_\infty + \sqrt{p^{-1}\log^5 n} \max_{i=1,2,3} \left\|\mathcal{M}_i^{\mathsf{T}}(\mathcal{Z})\right\|_{2,\infty}\right) \qquad (140)$$

holds with high probability.

Lemma 24 *Suppose $\mathcal{Z} \in \mathbb{R}^{n\times n\times n}$ is a tensor with multilinear rank (r_1, r_2, r_3). Then for any $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^{n\times 1}$ with $\|\boldsymbol{a}\|_2 = \|\boldsymbol{b}\|_2 = \|\boldsymbol{c}\|_2 = 1$, the tensor $\mathcal{Z} \odot (\boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c})$ has multilinear rank at most (r_1, r_2, r_3), where $\odot$ denotes the entrywise product.*

Proof [Proof of Lemma 24] Since $\mathcal{Z}$ is of multilinear rank (r_1, r_2, r_3), it can be decomposed as $\mathcal{Z} = \mathcal{H} \times_{i=1}^{3} \boldsymbol{Z}_i$, where $\mathcal{H} \in \mathbb{R}^{r_1\times r_2\times r_3}$ and $\boldsymbol{Z}_i \in \mathbb{R}^{n\times r_i}$ for $i = 1, 2, 3$. To complete the proof, it suffices to show that the matrix $\mathcal{M}_i(\mathcal{Z} \odot (\boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c}))$ is of rank at most r_i for $i = 1, 2, 3$. By the definition of entrywise product,

$$\begin{aligned}
\mathcal{M}_1(\mathcal{Z} \odot (\boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c})) &= \mathcal{M}_1(\mathcal{Z}) \odot \mathcal{M}_1(\boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c}) \\
&= \left(\boldsymbol{Z}_1\mathcal{M}_1(\mathcal{H})(\boldsymbol{Z}_3 \otimes \boldsymbol{Z}_2)^{\mathsf{T}}\right) \odot \left(\boldsymbol{a}(\boldsymbol{c} \otimes \boldsymbol{b})^{\mathsf{T}}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \left(\sum_{i=1}^{r_1} [\boldsymbol{Z}_1]_{:,i} \left[\mathcal{M}_1(\mathcal{H})(\boldsymbol{Z}_3 \otimes \boldsymbol{Z}_2)^\top\right]_{i,:}\right) \odot \left(\boldsymbol{a}(\boldsymbol{c} \otimes \boldsymbol{b})^\top\right) \\
&= \sum_{i=1}^{r_1} \left([\boldsymbol{Z}_1]_{:,i} \left[\mathcal{M}_1(\mathcal{H})(\boldsymbol{Z}_3 \otimes \boldsymbol{Z}_2)^\top\right]_{i,:}\right) \odot \left(\boldsymbol{a}(\boldsymbol{c} \otimes \boldsymbol{b})^\top\right) \\
&= \sum_{i=1}^{r_1} \left([\boldsymbol{Z}_1]_{:,i} \odot \boldsymbol{a}\right) \left(\left[\mathcal{M}_1(\mathcal{H})(\boldsymbol{Z}_3 \otimes \boldsymbol{Z}_2)^\top\right]_{i,:} \odot (\boldsymbol{c} \otimes \boldsymbol{b})^\top\right),
\end{aligned}
$$

which implies that this matrix has rank at most r_1. This is certainly true for $i = 2, 3$. ■

Lemma 25 (Uniform bound) *Suppose $\mathcal{J} \in \mathbb{R}^{n\times n\times n}$ is the full-one tensor. For any tensor $\mathcal{Z} \in \mathbb{R}^{n\times n\times n}$ with multilinear rank (r_1, r_2, r_3), the following inequality holds,*

$$
\left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{Z})\right\| \le \left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right\| \cdot \min\left\{\sqrt{r_1 r_2}, \sqrt{r_1 r_3}, \sqrt{r_2 r_3}\right\} \|\mathcal{Z}\|_\infty .
$$

Proof [Proof of Lemma 25] Let $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^{n\times 1}$ be unit vectors and $\tilde{r} = \min\left\{\sqrt{r_1 r_2}, \sqrt{r_1 r_3}, \sqrt{r_2 r_3}\right\}$. We have

$$
\begin{aligned}
\left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{Z})\right\| &= \left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J}) \odot \mathcal{Z}\right\| \\
&= \sup_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}} \left\langle \left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J}) \odot \mathcal{Z}, \boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c}\right\rangle \\
&= \sup_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}} \left\langle \left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J}), \mathcal{Z} \odot (\boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c})\right\rangle \\
&\le \left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right\| \cdot \sup_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}} \|\mathcal{Z} \odot (\boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c})\|_* \\
&\overset{(a)}{\le} \left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right\| \cdot \tilde{r} \sup_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}} \|\mathcal{Z} \odot (\boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c})\|_{\mathsf{F}} \\
&= \left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right\| \cdot \tilde{r} \sup_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}} \sqrt{\sum_{i_1,i_2,i_3} \mathcal{Z}_{i_1,i_2,i_3}^2 (a_{i_1} b_{i_2} c_{i_3})^2} \\
&\le \left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right\| \cdot \tilde{r} \sup_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}} \sqrt{\sum_{i_1,i_2,i_3} (a_{i_1} b_{i_2} c_{i_3})^2} \cdot \|\mathcal{Z}\|_\infty \\
&= \left\|\left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)(\mathcal{J})\right\| \cdot \tilde{r} \|\mathcal{Z}\|_\infty ,
\end{aligned}
$$

where (a) is due to Lemma 21 and Lemma 24. ■

Lemma 26 (Ma et al., 2020, Lemma 45; Chen et al., 2020a, Lemma 16) *Let $\boldsymbol{M}$ and $\widehat{\boldsymbol{M}} \in \mathbb{R}^{n\times n}$ be symmetric matrices with top-r eigenvalue decomposition $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ and $\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{U}}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{U}}$, respectively. Assume $\sigma_r(\boldsymbol{M}) > 0$, $\sigma_{r+1}(\boldsymbol{M}) = 0$ and $\left\|\boldsymbol{M} - \widehat{\boldsymbol{M}}\right\| \le \frac{1}{4}\sigma_r(\boldsymbol{M})$. Define $\boldsymbol{Q} = \arg\min_{\boldsymbol{R}^\top\boldsymbol{R}=\boldsymbol{I}} \left\|\widehat{\boldsymbol{U}}\boldsymbol{R} - \boldsymbol{U}\right\|_{\mathsf{F}}$. Then*

$$
\left\|\widehat{\boldsymbol{U}}\boldsymbol{Q} - \boldsymbol{U}\right\| \le \frac{3}{\sigma_r(\boldsymbol{M})} \left\|\boldsymbol{M} - \widehat{\boldsymbol{M}}\right\| .
$$

Lemma 27 *Let $\mathcal{T} \in \mathbb{R}^{n \times n \times n}$ be a tensor with Tucker decomposition $\mathcal{S} \times_{i=1}^{3} \boldsymbol{U}_i$ where $\boldsymbol{U}_i^\top \boldsymbol{U}_i = \boldsymbol{I} \in \mathbb{R}^{r \times r}$ for $i = 1, 2, 3$. Suppose $\mathcal{T}$ is μ-incoherent, and Ω obeys the Bernoulli observation with parameter p. If $p \geq C_2 \mu^2 r^2 \log n / n^2$, then with high probability,*

$$\left|\left\langle (\mathcal{I} - p^{-1}\mathcal{P}_\Omega)(\mathcal{X}),\ \mathcal{Z} \right\rangle\right| \leq C \sqrt{\frac{\mu^2 r^2 \log n}{n^2 p}} \, \|\mathcal{X}\|_{\mathsf{F}} \, \|\mathcal{Z}\|_{\mathsf{F}} \,,$$

holds simultaneously for all tensors $\mathcal{X}, \mathcal{Z} \in \mathbb{R}^{n \times n \times n}$ of the form

$$\mathcal{X} = \mathcal{G} \times_i \boldsymbol{X}_i \times_{j \neq i} \boldsymbol{U}_i, \mathcal{Z} = \mathcal{H} \times_i \boldsymbol{Z}_i \times_{j \neq i} \boldsymbol{U}_i,$$

where $\mathcal{G}, \mathcal{H} \in \mathbb{R}^{n \times r \times r}$ and $\boldsymbol{X}_i, \boldsymbol{Z}_i \in \mathbb{R}^{n \times n}$ are arbitrary factors.

Remark 28 *This lemma is essentially a restatement of Lemma 12 in Tong et al., 2021. Considering the case $i = 1$, the only difference is that here $\mathcal{G}$ (or $\mathcal{H}$) and $\boldsymbol{X}_i$ (or $\boldsymbol{Z}_i$) are respectively of size $n \times r \times r$ and $n \times n$ instead of $r \times r \times r$ and $n \times r$. However, the proof therein is equally applicable here and thus we omit the proof.*

Lemma 29 (Tong et al., 2021, Lemma 14) *Suppose Ω satisfies the Bernoulli observation model. Then with high probability,*

$$\left|\left\langle \left(\mathcal{I} - p^{-1}\mathcal{P}_\Omega\right)\left(\mathcal{G} \times_{i=1}^{3} \boldsymbol{X}_i\right), \mathcal{H} \times_{i=1}^{3} \boldsymbol{Z}_i \right\rangle\right| \leq C \left(p^{-1} \log^3 n + \sqrt{p^{-1} n \log^5 n}\right) \tau,$$

holds simultaneously for all tensors $\mathcal{G} \times_{i=1}^{3} \boldsymbol{X}_i$ and $\mathcal{H} \times_{i=1}^{3} \boldsymbol{Z}_i$, where the quantity τ obeys

$$\begin{aligned}\tau \leq &\left(\|\boldsymbol{X}_1 \mathcal{M}_1(\mathcal{G})\|_{2,\infty} \|\boldsymbol{Z}_1 \mathcal{M}_1(\mathcal{H})\|_{\mathsf{F}} \wedge \|\boldsymbol{X}_1 \mathcal{M}_1(\mathcal{G})\|_{\mathsf{F}} \|\boldsymbol{Z}_1 \mathcal{M}_1(\mathcal{H})\|_{2,\infty}\right) \\ &\cdot \left(\|\boldsymbol{X}_2\|_{2,\infty} \|\boldsymbol{Z}_2\|_{\mathsf{F}} \wedge \|\boldsymbol{X}_2\|_{\mathsf{F}} \|\boldsymbol{Z}_2\|_{2,\infty}\right) \left(\|\boldsymbol{X}_3\|_{2,\infty} \|\boldsymbol{Z}_3\|_{\mathsf{F}} \wedge \|\boldsymbol{X}_3\|_{\mathsf{F}} \|\boldsymbol{Z}_3\|_{2,\infty}\right).\end{aligned}$$

Lemma 30 *Suppose Ω satisfies the Bernoulli observation model. Then with high probability,*

$$\left|\left\langle \left(\mathcal{I} - p^{-1}\mathcal{P}_{\Omega_{-\ell}} - \mathcal{P}_\ell\right)\left(\mathcal{G} \times_{i=1}^{3} \boldsymbol{X}_i\right), \mathcal{H} \times_{i=1}^{3} \boldsymbol{Z}_i \right\rangle\right| \leq C \left(p^{-1} \log^3 n + \sqrt{p^{-1} n \log^5 n}\right) \tau,$$

holds simultaneously for all tensors $\mathcal{G} \times_{i=1}^{3} \boldsymbol{X}_i$ and $\mathcal{H} \times_{i=1}^{3} \boldsymbol{Z}_i$, where the quantity τ obeys

$$\begin{aligned}\tau \leq &\left(\|\boldsymbol{X}_1 \mathcal{M}_1(\mathcal{G})\|_{2,\infty} \|\boldsymbol{Z}_1 \mathcal{M}_1(\mathcal{H})\|_{\mathsf{F}} \wedge \|\boldsymbol{X}_1 \mathcal{M}_1(\mathcal{G})\|_{\mathsf{F}} \|\boldsymbol{Z}_1 \mathcal{M}_1(\mathcal{H})\|_{2,\infty}\right) \\ &\cdot \left(\|\boldsymbol{X}_2\|_{2,\infty} \|\boldsymbol{Z}_2\|_{\mathsf{F}} \wedge \|\boldsymbol{X}_2\|_{\mathsf{F}} \|\boldsymbol{Z}_2\|_{2,\infty}\right) \left(\|\boldsymbol{X}_3\|_{2,\infty} \|\boldsymbol{Z}_3\|_{\mathsf{F}} \wedge \|\boldsymbol{X}_3\|_{\mathsf{F}} \|\boldsymbol{Z}_3\|_{2,\infty}\right).\end{aligned}$$

Remark 31 *The proof of Lemma 30 is similar to the proof of Lemma 29, so we omit it.*

Lemma 32 (Ma et al., 2018, Lemma 37; Chen et al., 2020a, Lemma 6) *Suppose* $\boldsymbol{U}$, $\boldsymbol{X}_1$, $\boldsymbol{X}_2 \in \mathbb{R}^{n\times r}$ *are matrices such that*

$$\|\boldsymbol{X}_1 - \boldsymbol{U}\| \, \|\boldsymbol{U}\| \le \frac{\sigma_r^2(\boldsymbol{U})}{2} \quad \text{and} \quad \|\boldsymbol{X}_1 - \boldsymbol{X}_2\| \, \|\boldsymbol{U}\| \le \frac{\sigma_r^2(\boldsymbol{U})}{4}.$$

Let $\boldsymbol{R}_1$ *and* $\boldsymbol{R}_2$ *be orthogonal matrices such that*

$$\boldsymbol{R}_1 = \arg\min_{\boldsymbol{R}^\top\boldsymbol{R}=\boldsymbol{I}} \|\boldsymbol{X}_1\boldsymbol{R} - \boldsymbol{U}\|_{\mathsf{F}} \;\; \text{and} \;\; \boldsymbol{R}_2 = \arg\min_{\boldsymbol{R}^\top\boldsymbol{R}=\boldsymbol{I}} \|\boldsymbol{X}_2\boldsymbol{R} - \boldsymbol{U}\|_{\mathsf{F}}.$$

Then one has

$$\|\boldsymbol{X}_1\boldsymbol{R}_1 - \boldsymbol{X}_2\boldsymbol{R}_2\|_{\mathsf{F}} \le 5\frac{\sigma_1^2(\boldsymbol{U})}{\sigma_r^2(\boldsymbol{U})} \|\boldsymbol{X}_1 - \boldsymbol{X}_2\|_{\mathsf{F}}.$$

Lemma 33 (Wei et al., 2020, Lemma 4.1) *Let* $\widehat{\boldsymbol{X}} = \widehat{\boldsymbol{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{V}}^\top$ *and* $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ *be rank* r *matrices. Then*

$$\left\|\widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}^\top - \boldsymbol{U}\boldsymbol{U}^\top\right\| \le \frac{\left\|\widehat{\boldsymbol{X}} - \boldsymbol{X}\right\|_{\mathsf{F}}}{\sigma_{\min}(\boldsymbol{X})} \quad \text{and} \quad \left\|\widehat{\boldsymbol{V}}\widehat{\boldsymbol{V}}^\top - \boldsymbol{V}\boldsymbol{V}^\top\right\| \le \frac{\left\|\widehat{\boldsymbol{X}} - \boldsymbol{X}\right\|_{\mathsf{F}}}{\sigma_{\min}(\boldsymbol{X})}.$$

Lemma 34 (Cai and Zhang, 2018, Lemma 1) *Suppose* $\boldsymbol{V}, \widehat{\boldsymbol{V}} \in \mathbb{R}^{n\times r}$ *are orthonormal matrices. Then the following relations hold,*

$$\begin{aligned}
&\left\|\sin\Theta\left(\boldsymbol{V}, \widehat{\boldsymbol{V}}\right)\right\|_{\mathsf{F}} \le \inf_{\boldsymbol{R}^\top\boldsymbol{R}=\boldsymbol{I}} \left\|\widehat{\boldsymbol{V}} - \boldsymbol{V}\boldsymbol{R}\right\|_{\mathsf{F}} \le \sqrt{2}\left\|\sin\Theta\left(\boldsymbol{V}, \widehat{\boldsymbol{V}}\right)\right\|_{\mathsf{F}}, \\
&\left\|\sin\Theta\left(\boldsymbol{V}, \widehat{\boldsymbol{V}}\right)\right\| \le \inf_{\boldsymbol{R}^\top\boldsymbol{R}=\boldsymbol{I}} \left\|\widehat{\boldsymbol{V}} - \boldsymbol{V}\boldsymbol{R}\right\| \le \sqrt{2}\left\|\sin\Theta\left(\boldsymbol{V}, \widehat{\boldsymbol{V}}\right)\right\|, \\
&\left\|\sin\Theta\left(\boldsymbol{V}, \widehat{\boldsymbol{V}}\right)\right\| \le \left\|\widehat{\boldsymbol{V}}\widehat{\boldsymbol{V}}^\top - \boldsymbol{V}\boldsymbol{V}^\top\right\| \le 2\left\|\sin\Theta\left(\boldsymbol{V}, \widehat{\boldsymbol{V}}\right)\right\|, \\
&\left\|\widehat{\boldsymbol{V}}\widehat{\boldsymbol{V}}^\top - \boldsymbol{V}\boldsymbol{V}^\top\right\|_{\mathsf{F}} = \sqrt{2}\left\|\sin\Theta\left(\boldsymbol{V}, \widehat{\boldsymbol{V}}\right)\right\|_{\mathsf{F}}.
\end{aligned}$$

Lemma 35 (Davis-Kahan $\sin\Theta$ Theorem) *Suppose* $\boldsymbol{G}^\natural, \boldsymbol{\Delta} \in \mathbb{R}^{n\times n}$ *are symmetric matrices, and* $\widehat{\boldsymbol{G}} = \boldsymbol{G}^\natural + \boldsymbol{\Delta}$. *Let* $\delta = \sigma_r(\boldsymbol{G}^\natural) - \sigma_{r+1}(\boldsymbol{G}^\natural)$ *be the gap between the top* r*-th and* $r+1$*-th eigenvalues of* $\boldsymbol{G}^\natural$, *and* $\boldsymbol{U}$, $\widehat{\boldsymbol{U}}$ *be matrices whose columns are the top* r *orthonormal eigenvectors of* $\boldsymbol{G}^\natural$ *and* $\widehat{\boldsymbol{G}}$ *respectively. If* $\delta > \|\boldsymbol{\Delta}\|$, *then we have*

$$\left\|\sin\Theta\left(\boldsymbol{U}, \widehat{\boldsymbol{U}}\right)\right\| \le \frac{\|\boldsymbol{\Delta}\boldsymbol{U}\|}{\delta - \|\boldsymbol{\Delta}\|} \quad \text{and} \quad \left\|\sin\Theta\left(\boldsymbol{U}, \widehat{\boldsymbol{U}}\right)\right\|_{\mathsf{F}} \le \frac{\|\boldsymbol{\Delta}\boldsymbol{U}\|_{\mathsf{F}}}{\delta - \|\boldsymbol{\Delta}\|}.$$

Lemma 36 (Ding and Chen, 2020, Lemma 1) *Suppose* $\boldsymbol{G}^\natural, \boldsymbol{\Delta} \in \mathbb{R}^{n\times n}$ *are symmetric matrices, and* $\widehat{\boldsymbol{G}} = \boldsymbol{G}^\natural + \boldsymbol{\Delta}$. *The eigenvalue decomposition of* $\boldsymbol{G}^\natural$ *is denoted* $\boldsymbol{G}^\natural = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$, *where* $\boldsymbol{U} \in \mathbb{R}^{n\times r}$ *has orthonormal columns and* $\boldsymbol{\Lambda} = \mathrm{diag}(\sigma_1, \cdots, \sigma_r)$. *Let* $\boldsymbol{X} \in \mathbb{R}^{n\times r}$ *be the matrix whose columns are the top-*r *orthonormal eigenvectors of* $\widehat{\boldsymbol{G}}$. *Let the SVD of the matrix* $\boldsymbol{H} = \boldsymbol{X}^\top\boldsymbol{U}$ *be* $\boldsymbol{H} = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^\top$, *and define* $\boldsymbol{R} = \boldsymbol{A}\boldsymbol{B}^\top$. *If* $\|\boldsymbol{\Delta}\| < \frac{1}{2}\sigma_r$, *then one has*

$$\|\boldsymbol{\Lambda}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{\Lambda}\| \le \left(2 + \frac{\sigma_1}{\sigma_r - \|\boldsymbol{\Delta}\|}\right)\|\boldsymbol{\Delta}\|.$$

References

Anima Anandkumar, Dean P Foster, Daniel Hsu, Sham M Kakade, and Yi-Kai Liu. A spectral algorithm for latent dirichlet allocation. *Algorithmica*, 72(1):193–214, 2015.

Animashree Anandkumar, Rong Ge, Daniel Hsu, and Sham Kakade. A tensor spectral approach to learning mixed membership community models. In *Conference on Learning Theory*, pages 867–881. PMLR, 2013.

Richard Arratia and Louis Gordon. Tutorial on large deviations for the binomial distribution. *Bulletin of mathematical biology*, 51(1):125–131, 1989.

Boaz Barak and Ankur Moitra. Noisy tensor completion via the sum-of-squares hierarchy. In *Conference on Learning Theory*, pages 417–445. PMLR, 2016.

Changxiao Cai, Gen Li, Yuejie Chi, H Vincent Poor, and Yuxin Chen. Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *The Annals of Statistics*, 49(2):944–967, 2021a.

Changxiao Cai, Gen Li, H Vincent Poor, and Yuxin Chen. Nonconvex low-rank tensor completion from noisy data. *Operations Research*, 2021b.

Jian-Feng Cai, Lizhang Miao, Yang Wang, and Yin Xian. Provable near-optimal low-multilinear-rank tensor recovery. *arXiv preprint arXiv:2007.08904*, 2020.

Jian-Feng Cai, Jingyang Li, and Dong Xia. Provable tensor-train format tensor completion by riemannian optimization. *arXiv preprint arXiv:2108.12163*, 2021c.

T Tony Cai and Anru Zhang. Minimax rate-optimal estimation of high-dimensional covariance matrices with incomplete data. *Journal of multivariate analysis*, 150:55–74, 2016.

T Tony Cai and Anru Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46(1):60–89, 2018.

Han Chen, Garvesh Raskutti, and Ming Yuan. Non-convex projected gradient descent for generalized low-rank tensor regression. *The Journal of Machine Learning Research*, 20 (1):172–208, 2019.

Ji Chen, Dekai Liu, and Xiaodong Li. Nonconvex rectangular matrix completion via gradient descent without $\ell_{2,\infty}$ regularization. *IEEE Transactions on Information Theory*, 66(9): 5806–5841, 2020a.

Yuxin Chen, Yuejie Chi, Jianqing Fan, and Cong Ma. Spectral methods for data science: A statistical perspective. *arXiv preprint arXiv:2012.08496*, 2020b.

Juhee Cho, Donggyu Kim, and Karl Rohe. Asymptotic theory for estimating the singular vectors and values of a partially-observed low rank matrix with noise. *Statistica Sinica*, pages 1921–1948, 2017.

Andrzej Cichocki, Danilo Mandic, Lieven De Lathauwer, Guoxu Zhou, Qibin Zhao, Cesar Caiafa, and Huy Anh Phan. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE signal processing magazine*, 32(2): 145–163, 2015.

Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

Lijun Ding and Yudong Chen. Leave-one-out approach for matrix completion: Primal and dual analysis. *IEEE Transactions on Information Theory*, 66(11):7274–7301, 2020.

Andreas Elsener and Sara van de Geer. Sparse spectral estimation with missing and corrupted measurements. *Stat*, 8(1):e229, 2019.

Laura Florescu and Will Perkins. Spectral thresholds in the bipartite stochastic block model. In *Conference on Learning Theory*, pages 943–959. PMLR, 2016.

Silvia Gandy, Benjamin Recht, and Isao Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, 2011.

Gene H Golub and Charles F Van Loan. Matrix computations. johns hopkins studies in the mathematical sciences, 1996.

Rungang Han, Rebecca Willett, and Anru Zhang. An optimal statistical and computational framework for generalized tensor estimation. *arXiv preprint arXiv:2002.11255*, 2020.

Christopher J Hillar and Lek-Heng Lim. Most tensor problems are np-hard. *Journal of the ACM (JACM)*, 60(6):1–39, 2013.

Frank L Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927.

Bo Huang, Cun Mu, Donald Goldfarb, and John Wright. Provable models for robust low-rank tensor completion. *Pacific Journal of Optimization*, 11(2):339–364, 2015.

Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 665–674, 2013.

Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 79–86, 2010.

Raghunandan H Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE Transactions on Information Theory*, 56(6):2980–2998, 2010.

Othmar Koch and Christian Lubich. Dynamical tensor approximation. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2360–2375, 2010.

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

Daniel Kressner, Michael Steinlechner, and Bart Vandereycken. Low-rank tensor completion by riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, 2014.

Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):208–220, 2012.

Po-Ling Loh and Martin J Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics*, 40(3):1637–1664, 2012.

Karim Lounici. Sparse principal component analysis with missing observations. In *High dimensional probability VI*, pages 327–356. Springer, 2013.

Karim Lounici. High-dimensional covariance matrix estimation with missing observations. *Bernoulli*, 20(3):1029–1058, 2014.

Yuetian Luo and Anru R Zhang. Low-rank tensor estimation via riemannian gauss-newton: Statistical optimality and second-order convergence. *arXiv preprint arXiv:2104.12031*, 2021.

Cong Ma, Kaizheng Wang, Yuejie Chi, and Yuxin Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval and matrix completion. *Foundations of Computational Mathematics*, 2018.

Cong Ma, Kaizheng Wang, Yuejie Chi, and Yuxin Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. *Foundations of Computational Mathematics*, 20(3), 2020.

Andrea Montanari and Nike Sun. Spectral algorithms for tensor completion. *Communications on Pure and Applied Mathematics*, 71(11):2381–2425, 2018.

Cun Mu, Bo Huang, John Wright, and Donald Goldfarb. Square deal: Lower bounds and improved relaxations for tensor recovery. In *International conference on machine learning*, pages 73–81. PMLR, 2014.

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

Holger Rauhut, Reinhold Schneider, and Željka Stojanac. Low rank tensor recovery via iterative hard thresholding. *Linear Algebra and its Applications*, 523:220–262, 2017.

Benjamin Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12(12), 2011.

Tian Tong, Cong Ma, Ashley Prater-Bennette, Erin Tripp, and Yuejie Chi. Scaling and scalability: Provable nonconvex low-rank tensor estimation from incomplete measurements. *arXiv preprint arXiv:2104.14526*, 2021.

Joel A Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

Bart Vandereycken. Low-rank matrix completion by riemannian optimization. *SIAM Journal on Optimization*, 23(2):1214–1236, 2013.

Haifeng Wang, Jinchi Chen, and Ke Wei. Implicit regularization and entrywise convergence of riemannian optimization for low tucker-rank tensor completion. *arXiv preprint arXiv:2108.07899*, 2021.

Ke Wei, Jian-Feng Cai, Tony F Chan, and Shingyu Leung. Guarantees of Riemannian optimization for low rank matrix recovery. *SIAM Journal on Matrix Analysis and Applications*, 37(3):1198–1222, 2016.

Ke Wei, Jian-Feng Cai, Tony F Chan, and Shingyu Leung. Guarantees of Riemannian optimization for low rank matrix completion. *Inverse Problems & Imaging*, 14(2):233, 2020.

Dong Xia and Ming Yuan. On polynomial time methods for exact low-rank tensor completion. *Foundations of Computational Mathematics*, 19(6):1265–1313, 2019.

Dong Xia, Ming Yuan, and Cun-Hui Zhang. Statistically optimal and computationally efficient low rank tensor completion from noisy entries. *arXiv preprint arXiv:1711.04934*, 2017.

Dong Xia, Ming Yuan, and Cun-Hui Zhang. Statistically optimal and computationally efficient low rank tensor completion from noisy entries. *The Annals of Statistics*, 49(1), 2021.

Ming Yuan and Cun-Hui Zhang. On tensor completion via nuclear norm minimization. *Foundations of Computational Mathematics*, 16(4):1031–1068, 2016.

Anru R Zhang, T Tony Cai, and Yihong Wu. Heteroskedastic pca: Algorithm, optimality, and applications. *arXiv preprint arXiv:1810.08316*, 2018.

Zemin Zhang and Shuchin Aeron. Exact tensor completion using t-svd. *IEEE Transactions on Signal Processing*, 65(6):1511–1526, 2016.

Ziwei Zhu, Tengyao Wang, and Richard J Samworth. High-dimensional principal component analysis with heterogeneous missingness. *arXiv preprint arXiv:1906.12125*, 2019.