Fast and Provable Low-Rank High-Order Tensor Completion via Scaled Gradient Descent

Anonymous authors
Paper under double-blind review

Abstract

This work studies the low-rank high-order tensor completion problem, which aims to exactly recover a low-rank order-d ($d \geq 4$) tensor from partially observed entries. Recently, tensor Singular Value Decomposition (t-SVD)-based low-rank tensor completion has gained considerable attention due to its ability to capture the low-rank structure of multidimensional data. However, existing approaches often rely on the computationally expensive tensor nuclear norm (TNN), thereby limiting their scalability for real-world tensors. Leveraging the low-rank structure under the t-SVD decomposition, we propose an efficient algorithm that directly estimates the high-order tensor factors—starting from a spectral initialization—via scaled gradient descent (ScaledGD). Theoretically, we rigorously establish the recovery guarantees of the proposed algorithm under mild assumptions, demonstrating that it achieves linear convergence to the true low-rank tensor at a constant rate that is independent of the condition number. Numerical experiments on both synthetic and real-world data verify our results and demonstrate the superiority of our method.

1 Introduction

Tensors, or multidimensional arrays, are a natural representation of a wide range of real-world data, including color videos, medical images, tomographic images, hyperspectral images, audio data and beyond. Compared with representation using the vector/matrix structure, tensor provide a more powerful and flexible model for characterizing the intrinsic structural information underlying multidimensional data and multi-way interactions, and thus make it highly effective across a variety of applications, such as neuroscience (Ahmed et al., 2020), data mining (Papalexakis et al., 2016), signal processing (Sidiropoulos et al., 2017), and computer vision (Bibi & Ghanem, 2017; Zhang et al., 2021). Nevertheless, in many real-world applications, due to the defects caused by the signal acquisition process, such as occlusions and sensor failures, tensor data often suffer from information loss and noise corruptions. This motivates the two common tensor estimation problems, namely, tensor completion and tensor robust PCA. Tensor completion, which aims to recover the incomplete tensors from partial observations, has been widely studied in the literature; an incomplete list of works in this regard includes Liu et al. (2013); Goldfarb & Qin (2014); Zhang & Aeron (2017); Bengua et al. (2017); Lu et al. (2018); Zhou et al. (2018); Kong et al. (2018); Chen et al. (2019); Yuan et al. (2019); Lu et al. (2019); Jiang & Ng (2019); Huang et al. (2020); Song et al. (2020); Qin et al. (2022); Wang et al. (2022); Zhang & Ng (2022); Wang et al. (2023); Li et al. (2024); Qiu et al. (2024).

This problem is well studied in the matrix domain, where the goal is to exactly recover a low-rank matrix from an incomplete observation (Candès & Recht, 2009; Chen, 2015). However, the tensor extension of matrix completion is not an easy task because the tensor rank is not well defined. While there exist several different definitions of tensor rank, each of which has its own limitations. For example, the CANDECOMP/PARAFAC (CP) rank (Kiers, 2000), defined as the minimum number of factors in rank-one tensor decomposition, is generally NP-hard to compute and its convex relaxation is intractable (Hillar & Lim, 2013). As an alternative, the tractable Tucker rank (Tucker, 1966), is defined as a multilinear rank whose components are ranks of tensor matricization for all modes. Furthermore, as a convex surrogate for the Tucker rank, the sum-of-nuclear-norms (SNN) is defined as the sum of the nuclear norms of unfolding matrices (Liu et al., 2013). But it is still suboptimal since SNN is not the tightest convex relaxation of the Tucker rank (Romera-Paredes

& Pontil, 2013). There are some other definitions of tensor rank, we refer interested readers to Oseledets (2011) and Zhao et al. (2016) for further pointers.

All the aforementioned tensor decompositions model low-rankness in the original domain. Recently, an advanced tensor decomposition scheme called tensor singular value decomposition (t-SVD), induced by the notion of tensor-tensor product (t-product) (Kilmer et al., 2013), has received growing interests. The t-SVD factorizes a third-order tensor into the t-product of two orthogonal tensors and one f-diagonal tensor (also called singular value tensor). Accordingly, a new tensor rank called tensor tubal rank is defined as the number of nonzero singular tubes of the singular value tensor (Kilmer et al., 2013), and a new tensor nuclear norm has been proposed for low-tubal-rank approximation and applied for tensor completion (Zhang & Aeron, 2017; Lu et al., 2018; 2019; Jiang et al., 2020) and tensor robust PCA (Lu et al., 2020; Lu, 2021). The major advantage of t-SVD scheme over other tensor decomposition strategies in image processing applications is its capability to characterize the low-rank structures in the Fourier domain for tensors, especially for those tensors that have fixed orientation or certain spatial-shifting (Lu et al., 2020; Liu et al., 2020). In order to deal with high-order tensor data, such as color videos and remote sensing images, Qin et al. (2022) put forth a generic definition of high-order t-product based on any invertible linear transforms. Numerical examples have demonstrated its efficacy in tensor completion (Qin et al., 2022; Wang et al., 2023). However, one major shortcoming of these methods is that they rely on TNN and involve full t-SVD computation in each iteration, thus the resulting optimization programs are computationally rather expensive to solve, even for medium size tensors.

Our goal in this paper is to perform tensor completion based on the high-order t-SVD algebraic framework. Motivated by the recent success of scaled gradient descent (ScaledGD) (Tong et al., 2021; 2022; Wu, 2025), we propose a scaled gradient descent (ScaledGD) method for high-order tensor completion while maintaining the low per-iteration computational complexity. Specifically, it first factorizes the low-rank tensor as the high-order t-product of two factors with smaller sizes. These two factors are then updated by the scaled gradient descent algorithm that avoids performing high-order t-SVD on a full-sized tensor. Theoretically, we establish that ScaledGD can achieve linear convergence at a rate independent of the condition number of the ground truth tensor, as long as the sample complexity is large enough. Experiments on both synthetic and real-world data show the efficiency and effectiveness of the proposed algorithm.

Outline. Section 2 summarizes related work. Section 3 introduces the fundamental high-order t-SVD framework. The main model and theory are given in Section 4. We present experimental results in Section 5 and provide concluding remarks in Section 6. All proof details are given in the Appendix.

2 Related Work

Significant efforts have been devoted to understanding nonconvex optimization for low-rank matrix completion in recent years (Mishra et al., 2012; Tanner & Wei, 2016). To deal with multi-way data, while one could reshape the tensor into a large-scale matrix and apply matrix completion algorithms, such a preprocessing inevitably breaks the high-order information of the original tensor and can degrade the recovery performance. Common tensor decompositions include CP (Kolda & Bader, 2009), Tucker (Tucker, 1966), HOSVD (Lathauwer et al., 2000), t-SVD Kilmer et al. (2013); Zhang et al. (2014); Qin et al. (2022), and tensor networks (e.g., tensor tree, tensor train, tensor ring) (Oseledets, 2011; Ballani & Grasedyck, 2014; Zhao et al., 2016). Our work rests on the (high-order) t-SVD, which has proven to be highly effective in various image/video processing applications because it can exploit the low-rank structure in the frequency domain. Under the t-SVD framework, theoretical guarantee for the exact recovery has been provided for the third-order tensor completion problem in Zhang & Aeron (2017); Lu et al. (2018), which has been further extended by replacing the Discrete Fourier Transform (DFT) conducted along the third dimension by general choices of the invertible linear transforms (Lu et al., 2019). For high-order tensor completion, a low t-SVD rank tensor completion model with theoretical guarantee under the order-d t-SVD framework has been studied in Qin et al. (2022). While these approaches have been incredibly successful in many applications, an important shortcoming is that they are not scalable to large-scale tensor data due to the heavy t-SVD computational overhead required in each iteration.

To alleviate this issue, decomposition-based approaches have gained considerable attention, which depict the low-rank structure of a tensor by factorizing it into the t-product of two smaller tensors. Motivated by TNN, a low-rank tensor factorization method is proposed in Zhou et al. (2018) for solving the third-order tensor completion problem, and the proposed alternating minimization algorithm is proved to converge to a Karush-Kuhn-Tucker (KKT) point. Recently, Wu (2025) extended ScaledGD (Tong et al., 2021) to the low-rank tensor estimation problem under the t-SVD decomposition and provided the theoretical recovery guarantee. These approaches are only relevant to third-order tensors. Our work can be considered an extension of Wu (2025) to the high-order tensor completion problem.

3 Notations and Preliminaries

In this paper, we use bold calligraphic letters for tensors, e.g., $\boldsymbol{\mathcal{A}}$, bold uppercase letters for matrices, e.g., $\boldsymbol{A}$, bold lowercase letters for vectors, e.g., $\boldsymbol{a}$, and non-bold letters for scalars, e.g., a. Throughout this paper, the fields of real number and complex number are denoted as $\mathbb{R}$ and $\mathbb{C}$, respectively. For an order-d tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, its $(i_1, i_2, \ldots, i_d)$-th element is represented as $\boldsymbol{\mathcal{A}}_{i_1,\ldots,i_d}$. We denote the horizontal slices of $\boldsymbol{\mathcal{A}}$ as $\boldsymbol{\mathcal{A}}(i_1, :, \ldots, :)$. We denote the $(i_3, \ldots, i_d)$-th frontal slice of $\boldsymbol{\mathcal{A}}$ as $\boldsymbol{\mathcal{A}}(:, :, i_3, \ldots, i_d)$, which is also written as $\boldsymbol{A}^{(i_3,\ldots,i_d)}$. All these frontal slices can be indexed using a single index j, where the j-th frontal slice $\boldsymbol{A}^j$, with $j = (i_d - 1)n_3 \ldots n_{d-1} + \cdots + (i_4 - 1)n_3 + i_3$, corresponds to $\boldsymbol{A}^{(i_3,\ldots,i_d)}$. Then $\texttt{bdiag}(\boldsymbol{\mathcal{A}}) = \texttt{diag}(\boldsymbol{A}^1, \boldsymbol{A}^2, \ldots, \boldsymbol{A}^{J-1}, \boldsymbol{A}^J) \in \mathbb{R}^{n_1 n_3 \ldots n_d \times n_2 n_3 \ldots n_d}$, where $J = n_3 \cdots n_d$, is the block diagonal matrix constructed by all frontal slices.

The Frobenius and infinity norms of a tensor $\boldsymbol{\mathcal{A}}$ are defined as $\|\boldsymbol{\mathcal{A}}\|_F = \sqrt{\sum_{i_1,\ldots,i_d} |\boldsymbol{\mathcal{A}}_{i_1,\ldots,i_d}|^2}$ and $\|\boldsymbol{\mathcal{A}}\|_\infty = \max_{i_1,\ldots,i_d} |\boldsymbol{\mathcal{A}}_{i_1,\ldots,i_d}|$, respectively. The spectral norm of a matrix $\boldsymbol{A}$ is denoted as $\|\boldsymbol{A}\| = \max_i \sigma_i(\boldsymbol{A})$, where $\sigma_i(\boldsymbol{A})$'s are the singular values of $\boldsymbol{A}$. In particular, we use $\sigma_{\min}$ and $\sigma_{\max}$ to denote the minimum and maximum singular value of a matrix, respectively. The matrix nuclear norm of $\boldsymbol{A}$ is $\|\boldsymbol{A}\|_* = \sum_i \sigma_i(\boldsymbol{A})$. It is often convenient to unfold a tensor into a matrix. A mode-k unfolding of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is defined as $\texttt{unfold}_k(\boldsymbol{\mathcal{A}}) = \boldsymbol{\mathcal{A}}_{(k)} \in \mathbb{R}^{n_k \times (n_1 \ldots n_{k-1} n_{k+1} \ldots n_d)}$, and the reverse of this process is called mode-k folding, such that $\texttt{fold}_k(\boldsymbol{\mathcal{A}}_{(k)}) = \boldsymbol{\mathcal{A}}$. The conjugate transpose of a matrix $\boldsymbol{A} \in \mathbb{C}^{n_1 \times n_2}$ is denoted by $\boldsymbol{A}^*$. The $n \times n$ identity matrix is denoted by $\boldsymbol{I}_n$. We denote $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$, and $[a] \coloneqq \{1, 2, \ldots, a\}$. Further, $f(n) \gtrsim g(n)$ (resp., $f(n) \lesssim g(n)$) means $|f(n)|/|g(n)| \geq c$ (resp., $|f(n)|/|g(n)| \leq c$) for some constant $c > 0$ when n is sufficiently large.

Definition 1 (mode-k product (Kolda & Bader, 2009))**.** The mode-k product of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with a matrix $\boldsymbol{M} \in \mathbb{R}^{p \times n_k}$ is a tensor whose mode-k unfolding is $\boldsymbol{M}$ multiplied with the mode-k unfolding of $\boldsymbol{\mathcal{A}}$, i.e.,

$$\boldsymbol{\mathcal{A}} \times_k \boldsymbol{M} \coloneqq \texttt{fold}(\boldsymbol{M}\boldsymbol{\mathcal{A}}_{(k)}),$$

where $\texttt{fold}(\boldsymbol{M}\boldsymbol{\mathcal{A}}_{(k)}) \in \mathbb{R}^{n_1 \times \cdots \times n_{k-1} \times p \times n_{k+1} \times \cdots \times n_d}$.

Definition 2 (Facewise product (Qin et al., 2022))**.** The facewise product of two order-d tensors $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times p \times n_3 \times \cdots \times n_d}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{p \times n_2 \times n_3 \times \cdots \times n_d}$ is a tensor $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} \triangle \boldsymbol{\mathcal{B}} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots \times n_d}$, where each frontal slice of $\boldsymbol{\mathcal{C}}$ is the matrix multiplication of the corresponding frontal slices of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$, i.e.,

$$\boldsymbol{C}^{(i_3,\ldots,i_d)} = \boldsymbol{A}^{(i_3,\ldots,i_d)} \boldsymbol{B}^{(i_3,\ldots,i_d)}$$

for $i_k \in [n_k]$, $k = 3, \ldots, d$.

Before introducing the order-d t-product, we define the linear transform $\mathfrak{L}(\cdot) : \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \to \mathbb{C}^{n_1 \times n_2 \times \cdots \times n_d}$ associated with a set of invertible matrices $\{\boldsymbol{U}_k \in \mathbb{C}^{n_k \times n_k}\}_{k=3}^d$ with inverse mapping $\mathfrak{L}^{-1}(\cdot)$ as $\boldsymbol{\mathcal{A}}_{\mathfrak{L}} \coloneqq \mathfrak{L}(\boldsymbol{\mathcal{A}}) = \boldsymbol{\mathcal{A}} \times_3 \boldsymbol{U}_{n_3} \times \cdots \times_d \boldsymbol{U}_{n_d}$ and $\mathfrak{L}^{-1}(\boldsymbol{\mathcal{A}}) \coloneqq \boldsymbol{\mathcal{A}} \times_d \boldsymbol{U}_{n_d}^{-1} \times \cdots \times_3 \boldsymbol{U}_{n_3}^{-1}$ satisfying $\mathfrak{L}^{-1}(\mathfrak{L}(\boldsymbol{\mathcal{A}})) = \boldsymbol{\mathcal{A}}$. In this work, the transform matrices $\{\boldsymbol{U}_{n_k}\}_{k=3}^d$ of $\mathfrak{L}$ are assumed to satisfy

$$\begin{aligned}
&(\boldsymbol{U}_{n_d}^* \otimes \boldsymbol{U}_{n_{d-1}}^* \otimes \cdots \otimes \boldsymbol{U}_{n_3}^*) \cdot (\boldsymbol{U}_{n_d} \otimes \boldsymbol{U}_{n_{d-1}} \otimes \cdots \otimes \boldsymbol{U}_{n_3}) \\
&= (\boldsymbol{U}_{n_d} \otimes \boldsymbol{U}_{n_{d-1}} \otimes \cdots \otimes \boldsymbol{U}_{n_3}) \cdot (\boldsymbol{U}_{n_d}^* \otimes \boldsymbol{U}_{n_{d-1}}^* \otimes \cdots \otimes \boldsymbol{U}_{n_3}^*) \\
&= \ell \boldsymbol{I}_{n_3 \ldots n_d},
\end{aligned} \tag{1}$$

where $\otimes$ denotes the Kronecker product and $\ell > 0$ is specific scale factor corresponding to the transform. For example, $\ell = n_3 \cdots n_d$ for discrete fourier transform (DFT) matrices, since $\boldsymbol{U}_{n_k}^* \boldsymbol{U}_{n_k} = n_k \boldsymbol{I}_{n_k}$, and $\ell = 1$ for discrete cosine transform (DCT) matrices, since $\boldsymbol{U}_{n_k}^* \boldsymbol{U}_{n_k} = \boldsymbol{I}_{n_k}$, $k = 3, \ldots, d$.

Definition 3 (t-product (Qin et al., 2022))**.** Let $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times p \times n_3 \times \cdots \times n_d}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{p \times n_2 \times n_3 \times \cdots \times n_d}$, then the t-product $\boldsymbol{\mathcal{C}} \coloneqq \boldsymbol{\mathcal{A}} *_{\mathfrak{L}} \boldsymbol{\mathcal{B}}$ under the invertible transform $\mathfrak{L}$ is defined as

$$\boldsymbol{\mathcal{C}} \coloneqq \boldsymbol{\mathcal{A}} *_{\mathfrak{L}} \boldsymbol{\mathcal{B}} = \mathfrak{L}^{-1}(\mathfrak{L}(\boldsymbol{\mathcal{A}}) \triangle \mathfrak{L}(\boldsymbol{\mathcal{B}})). \tag{2}$$

Definition 4 (Conjugate transpose (Qin et al., 2022))**.** The conjugate transpose of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{n_1 \times n_2 \times n_3 \times \cdots \times n_d}$ is the tensor $\boldsymbol{\mathcal{A}}^T \in \mathbb{C}^{n_2 \times n_1 \times n_3 \times \cdots \times n_d}$ such that $\boldsymbol{\mathcal{A}}^T(:,:,i_3,\ldots,i_d) = (\boldsymbol{\mathcal{A}}(:,:,i_3,\ldots,i_d))^*$ for all $i_k \in [n_k]$, $k = 3, \ldots, d$.

Definition 5 (Identity tensor (Qin et al., 2022))**.** The order-d identity tensor $\boldsymbol{\mathcal{I}}_n \in \mathbb{R}^{n \times n \times n_3 \times \cdots \times n_d}$ is the tensor such that $\boldsymbol{\mathcal{I}}_{\mathfrak{L}}(:,:,i_3,\ldots,i_d) = \boldsymbol{I}_n$ for $i_k \in [n_k]$, $k = 3, \ldots, d$.

Definition 6 (Orthogonal tensor (Qin et al., 2022))**.** A tensor $\boldsymbol{\mathcal{Q}} \in \mathbb{C}^{n \times n \times n_3 \times \cdots \times n_d}$ is orthogonal if $\boldsymbol{\mathcal{Q}} *_{\mathfrak{L}} \boldsymbol{\mathcal{Q}}^T = \boldsymbol{\mathcal{Q}}^T *_{\mathfrak{L}} \boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{I}}_n$.

Definition 7 (f-diagonal tensor (Qin et al., 2022))**.** A tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is called f-diagonal if all the frontal slices of $\boldsymbol{\mathcal{A}}$ are diagonal matrices.

Definition 8 (Tensor inverse)**.** For an arbitrary tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{r \times r \times n_3 \times \cdots \times n_d}$, its inverse is defined as a tensor $\boldsymbol{\mathcal{A}}^{-1} \in \mathbb{R}^{r \times r \times n_3 \times \cdots \times n_d}$ which satisfies $\boldsymbol{\mathcal{A}} *_{\mathfrak{L}} \boldsymbol{\mathcal{A}}^{-1} = \boldsymbol{\mathcal{A}}^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{I}}_r$. The set of invertible tensors in $\mathbb{R}^{r \times r \times n_3 \times \cdots \times n_d}$ is denoted by $\mathrm{GL}(r)$.

Definition 9 (Tensor $\ell_{2,\infty}$-norm)**.** The tensor $\ell_{2,\infty}$-norm of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is defined as $\|\boldsymbol{\mathcal{A}}\|_{2,\infty} = \max_{i_1} \|\boldsymbol{\mathcal{A}}(i_1,:,:,\ldots,:)\|_F$.

Theorem 1 (tensor singular value decomposition (t-SVD) (Qin et al., 2022))**.** *For any order-d tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, it can be decomposed as*

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}} *_{\mathfrak{L}} \boldsymbol{\mathcal{V}}^T, \tag{3}$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n_1 \times n_1 \times n_3 \times \cdots \times n_d}$ and $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{n_2 \times n_2 \times n_3 \times \cdots \times n_d}$ are orthogonal tensors, and $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots \times n_d}$ is a rectangular f-diagonal tensor.

One can obtain t-SVD efficiently by performing matrix SVDs in the transformed domain as shown in Algorithm 1. The entries on the diagonal of the first frontal slice $\boldsymbol{\mathcal{S}}(:,:,1,\ldots,1)$ of $\boldsymbol{\mathcal{S}}$ have the decreasing property, i.e., $\boldsymbol{\mathcal{S}}_{1,1,1,\ldots,1} \geq \boldsymbol{\mathcal{S}}_{2,2,1,\ldots,1} \geq \cdots \geq \boldsymbol{\mathcal{S}}_{n',n',1,\ldots,1}$, where $n' = \min\{n_1, n_2\}$. The diagonal entries of $\boldsymbol{\mathcal{S}}_{\mathfrak{L}}(:,:,i_3,\ldots,i_d)$ correspond to the singular values of $\boldsymbol{\mathcal{A}}_{\mathfrak{L}}(:,:,i_3,\ldots,i_d)$.

Definition 10 (Tensor multi-rank (Qin et al., 2022))**.** The multi-rank of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with respect to the invertible transforms $\mathfrak{L}$ is a vector $\boldsymbol{r} \in \mathbb{R}^{n_3 \cdots n_d}$, in which the i-th element of $\boldsymbol{r}$ equals to the rank of the i-th block of $\mathtt{bdiag}(\boldsymbol{\mathcal{A}}_{\mathfrak{L}})$.

Definition 11 (t-SVD rank (Qin et al., 2022))**.** Let $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}} *_{\mathfrak{L}} \boldsymbol{\mathcal{V}}^T$ be the t-SVD of $\boldsymbol{\mathcal{A}}$, its t-SVD rank $\mathrm{rank}_{\text{t-SVD}}(\boldsymbol{\mathcal{A}})$ is defined as

$$\mathrm{rank}_{\text{t-SVD}}(\boldsymbol{\mathcal{A}}) = \#\{i : \boldsymbol{\mathcal{S}}(i,i,:,\ldots,:) \neq \boldsymbol{0}\},$$

where $\#$ denotes the cardinality of a set.

Definition 12 (tensor nuclear norm (Qin et al., 2022))**.** Let $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, the tensor nuclear norm of $\boldsymbol{\mathcal{A}}$ is defined as

$$\|\boldsymbol{\mathcal{A}}\|_{\circledast,\mathfrak{L}} \coloneqq \frac{1}{\ell} \sum_{i_3=1}^{n_3} \cdots \sum_{i_d=1}^{n_d} \|\boldsymbol{\mathcal{A}}_{\mathfrak{L}}(:,:,i_3,\ldots,i_d)\|_*,$$

where $\|\cdot\|_*$ denotes the nuclear norm of a matrix.

4 Main Results

In this section, we introduce ScaledGD for high-order tensor completion and establish its performance guarantee.

Algorithm 1 t-SVD for order-d tensors (Qin et al., 2022)

Input: $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ and the corresponding matrices $\{\boldsymbol{U}_{n_k}\}_{k=3}^{d}$ of invertible transform $\mathfrak{L}$.

1: $\boldsymbol{\mathcal{A}}_{\mathfrak{L}} = \mathfrak{L}(\boldsymbol{\mathcal{A}})$.
2: Compute each frontal slice of $\boldsymbol{\mathcal{U}}_{\mathfrak{L}}$, $\boldsymbol{\mathcal{S}}_{\mathfrak{L}}$ and $\boldsymbol{\mathcal{V}}_{\mathfrak{L}}$ from $\boldsymbol{\mathcal{A}}_{\mathfrak{L}}$ by
3: **for** $i_3 \in [n_3], \ldots, i_d \in [n_d]$ **do**
4: $\quad [\boldsymbol{U}, \boldsymbol{S}, \boldsymbol{V}] = \mathrm{SVD}(\boldsymbol{\mathcal{A}}_{\mathfrak{L}}(:,:,i_3,\ldots,i_d))$,
5: $\quad \boldsymbol{\mathcal{U}}_{\mathfrak{L}}(:,:,i_3,\ldots,i_d) = \boldsymbol{U}$, $\boldsymbol{\mathcal{S}}_{\mathfrak{L}}(:,:,i_3,\ldots,i_d) = \boldsymbol{S}$, $\boldsymbol{\mathcal{V}}_{\mathfrak{L}}(:,:,i_3,\ldots,i_d) = \boldsymbol{V}$.
6: **end for**
7: $\boldsymbol{\mathcal{U}} = \mathfrak{L}^{-1}(\boldsymbol{\mathcal{U}}_{\mathfrak{L}})$, $\boldsymbol{\mathcal{S}} = \mathfrak{L}^{-1}(\boldsymbol{\mathcal{S}}_{\mathfrak{L}})$, $\boldsymbol{\mathcal{V}} = \mathfrak{L}^{-1}(\boldsymbol{\mathcal{V}}_{\mathfrak{L}})$.

Output: t-SVD components $\boldsymbol{\mathcal{U}}$, $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{V}}$ such that $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}} *_{\mathfrak{L}} \boldsymbol{\mathcal{V}}^T$.

4.1 Problem Formulation

Suppose that the ground truth t-SVD rank-r tensor $\boldsymbol{\mathcal{X}}_\star \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ admits the following compact t-SVD decomposition $\boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{U}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{V}}_\star^T$, where $\boldsymbol{\mathcal{U}}_\star \in \mathbb{R}^{n_1 \times r \times n_3 \times \cdots \times n_d}$, $\boldsymbol{\mathcal{V}}_\star \in \mathbb{R}^{n_2 \times r \times n_3 \times \cdots \times n_d}$, and $\boldsymbol{\mathcal{S}}_\star \in \mathbb{R}^{r \times r \times n_3 \times \cdots \times n_d}$. We define its top-$r$ tensor factors as

$$\boldsymbol{\mathcal{L}}_\star = \boldsymbol{\mathcal{U}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} \quad \text{and} \quad \boldsymbol{\mathcal{R}}_\star = \boldsymbol{\mathcal{V}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}$$

so that $\boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T$. Here, the "square root" of a tensor $\boldsymbol{\mathcal{A}}$, denoted by $\boldsymbol{\mathcal{A}}^{\frac{1}{2}}$, is obtained by setting $\boldsymbol{\mathcal{A}}^{\frac{1}{2}} := \boldsymbol{\mathcal{B}} = L^{-1}(\boldsymbol{\mathcal{B}}_{\mathfrak{L}})$, where the $(i_3, \ldots, i_d)$-th frontal slice of $\boldsymbol{\mathcal{B}}_{\mathfrak{L}}$ as $\boldsymbol{\mathcal{B}}_{\mathfrak{L}}(:,:,i_3,\ldots,i_d) = (\boldsymbol{\mathcal{A}}_{\mathfrak{L}}(:,:,i_3,\ldots,i_d))^{\frac{1}{2}}$ for all $i_k \in [n_k]$, $k = 3, \ldots, d$.

Assume that we have observed a subset $\boldsymbol{\Omega}$ of the entries of $\boldsymbol{\mathcal{X}}_\star$, denoted by $\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star)$, where $\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\cdot)$ represents a linear operator such that

$$[\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}})]_{i_1,\ldots,i_d} = \begin{cases} \boldsymbol{\mathcal{X}}_{i_1,\ldots,i_d}, & \text{if } (i_1,\ldots,i_d) \in \boldsymbol{\Omega}, \\ 0, & \text{otherwise.} \end{cases}$$

Here, $\boldsymbol{\Omega}$ is generated according to the Bernoulli model in the sense that $\boldsymbol{\Omega} = \{(i_1,\ldots,i_d) | \delta_{i_1 \ldots i_d} = 1\}$, where $\delta_{i_1 \ldots i_d}$'s are independent and identically distributed (i.i.d.) variables taking value one with probability p. In this setting, we denote that $\boldsymbol{\Omega} \sim \mathrm{Ber}(p)$. The problem of tensor completion is to recover the underlying low t-SVD rank tensor $\boldsymbol{\mathcal{X}}$ from the partial observations $\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star)$. To estimate $\boldsymbol{\mathcal{X}}_\star$ more efficiently, we parameterize $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{L}} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}^T$ by two low-rank factors $\boldsymbol{\mathcal{L}} \in \mathbb{R}^{n_1 \times r \times n_3 \times \cdots \times n_d}$ and $\boldsymbol{\mathcal{R}} \in \mathbb{R}^{n_2 \times r \times n_3 \times \cdots \times n_d}$ as in Tong et al. (2021); Wu (2025) and solve the following optimization problem:

$$\min_{\substack{\boldsymbol{\mathcal{L}} \in \mathbb{R}^{n_1 \times r \times n_3 \times \cdots \times n_d} \\ \boldsymbol{\mathcal{R}} \in \mathbb{R}^{n_2 \times r \times n_3 \times \cdots \times n_d}}} f(\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{R}}) = \frac{1}{2p} \|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}^T - \boldsymbol{\mathcal{X}}_\star)\|_F^2. \tag{4}$$

Obviously, the tensor completion problem is ill-posed without imposing additional constraints on the low-rank tensor $\boldsymbol{\mathcal{X}}_\star$, which are crucial in determining the performance of the proposed algorithm. We first introduce the incoherence parameter of the tensor $\boldsymbol{\mathcal{X}}_\star$.

Definition 13 (Incoherence)**.** For $\boldsymbol{\mathcal{X}}_\star \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with t-SVD rank r, assume that it has the skinny t-SVD $\boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{U}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{V}}_\star^T$. Then $\boldsymbol{\mathcal{X}}_\star$ is said to satisfy the tensor incoherence conditions with parameter μ if

$$\max_{i_1 \in [n_1]} \|\boldsymbol{\mathcal{U}}_\star^T *_{\mathfrak{L}} \mathring{\mathbf{e}}_1^{(i_1)}\|_F \le \sqrt{\frac{\mu r}{n_1 \ell}} \quad \text{and} \quad \max_{i_2 \in [n_2]} \|\boldsymbol{\mathcal{V}}_\star^T *_{\mathfrak{L}} \mathring{\mathbf{e}}_2^{(i_2)}\|_F \le \sqrt{\frac{\mu r}{n_2 \ell}}. \tag{5}$$

Here, $\mathring{\mathbf{e}}_1^{(i_1)}$ is the order-d tensor mode-1 basis of size $n_1 \times 1 \times n_3 \times \cdots \times n_d$ with its $(i_1, 1, 1, \ldots, 1)$-th entry equaling to 1 and the rest equaling to 0, and $\mathring{\mathbf{e}}_2^{(i_2)} := (\mathring{\mathbf{e}}_1^{(i_2)})^T$ is the mode-2 basis.

Next, for $\boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{U}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{V}}_\star^T$ with multi-rank $\boldsymbol{r}$, we define the following two important singular values of tensor $\boldsymbol{\mathcal{X}}_\star$ as

$$\begin{aligned} &\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star) = \max\{[\boldsymbol{\mathcal{S}}_{\mathfrak{L}}]_{i,i,i_3,\dots,i_d} | [\boldsymbol{\mathcal{S}}_{\mathfrak{L}}]_{i,i,i_3,\dots,i_d} > 0, i \le \min\{n_1, n_2\}, i_k \in [n_k], k = 3, \dots, d\} \\ \text{and} \quad &\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) = \min\{[\boldsymbol{\mathcal{S}}_{\mathfrak{L}}]_{i,i,i_3,\dots,i_d} | [\boldsymbol{\mathcal{S}}_{\mathfrak{L}}]_{i,i,i_3,\dots,i_d} > 0, i \le \min\{n_1, n_2\}, i_k \in [n_k], k = 3, \dots, d\}, \end{aligned}$$

where $s_r = \sum_{k=1}^{n_3 \cdots n_d} r_k$. Then the condition number κ of $\boldsymbol{\mathcal{X}}_\star$ is defined as

$$\kappa \coloneqq \bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star)/\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).$$

4.2 Proposed Algorithm

To minimize (4), our ScaledGD algorithm consists of two parts: (a) Spectral initialization and (b) Scaled gradient updates. We start with initializing $\boldsymbol{\mathcal{L}}$ and $\boldsymbol{\mathcal{R}}$ by setting $\boldsymbol{\mathcal{L}}_0 = \boldsymbol{\mathcal{U}}_0 *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_0^{\frac{1}{2}}$ and $\boldsymbol{\mathcal{R}}_0 = \boldsymbol{\mathcal{V}}_0 *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_0^{\frac{1}{2}}$, where $\boldsymbol{\mathcal{U}}_0 *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_0 *_{\mathfrak{L}} \boldsymbol{\mathcal{V}}_0^T$ is the best t-SVD rank-r approximation of $\frac{1}{p}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star)$. Next, we update the tensor factors iteratively along the scaled gradient directions:

$$\begin{aligned} \boldsymbol{\mathcal{L}}_{t+1} &= \boldsymbol{\mathcal{L}}_t - \eta \nabla_{\boldsymbol{\mathcal{L}}} f(\boldsymbol{\mathcal{L}}_t, \boldsymbol{\mathcal{R}}_t) *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_t^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t)^{-1}, \\ \boldsymbol{\mathcal{R}}_{t+1} &= \boldsymbol{\mathcal{R}}_t - \eta \nabla_{\boldsymbol{\mathcal{R}}} f(\boldsymbol{\mathcal{L}}_t, \boldsymbol{\mathcal{R}}_t) *_{\mathfrak{L}} (\boldsymbol{\mathcal{L}}_t^T *_{\mathfrak{L}} \boldsymbol{\mathcal{L}}_t)^{-1}, \end{aligned} \tag{6}$$

where $\eta > 0$ is the learning rate and $\nabla_{\boldsymbol{\mathcal{L}}} f(\boldsymbol{\mathcal{L}}_t, \boldsymbol{\mathcal{R}}_t)$ (resp., $\nabla_{\boldsymbol{\mathcal{R}}} f(\boldsymbol{\mathcal{L}}_t, \boldsymbol{\mathcal{R}}_t)$) denotes the gradient of f with respect to $\boldsymbol{\mathcal{L}}_t$ (resp., $\boldsymbol{\mathcal{R}}_t$) at the t-th iteration. To guarantee good performance from partial observations, the underlying low-rank tensor $\boldsymbol{\mathcal{X}}_\star$ needs to be incoherent (cf. Definition 13) to avoid ill-posedness. Similar to the matrix case (Chen & Wainwright, 2015), we trim the all the horizontal slices of $\boldsymbol{\mathcal{L}}$ and $\boldsymbol{\mathcal{R}}$ after the gradient update. To be specific, we introduce the scaled projection as follows (Tong et al., 2021; Wu, 2025):

$$\mathcal{P}_B\Big(\begin{bmatrix} \widetilde{\boldsymbol{\mathcal{L}}} \\ \widetilde{\boldsymbol{\mathcal{R}}} \end{bmatrix}\Big) = \begin{bmatrix} \boldsymbol{\mathcal{L}} \\ \boldsymbol{\mathcal{R}} \end{bmatrix},$$

where $B > 0$ is the projection radius, and

$$\begin{aligned} \boldsymbol{\mathcal{L}}(i,:,\dots,:) &= \Big(1 \wedge \frac{B}{\sqrt{n_1}\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\dots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{R}}}^T\|_F}\Big)\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\dots,:), \quad i \in [n_1], \\ \text{and} \quad \boldsymbol{\mathcal{R}}(j,:,\dots,:) &= \Big(1 \wedge \frac{B}{\sqrt{n_2}\|\widetilde{\boldsymbol{\mathcal{R}}}(j,:,\dots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{L}}}^T\|_F}\Big)\widetilde{\boldsymbol{\mathcal{R}}}(j,:,\dots,:), \quad j \in [n_2]. \end{aligned} \tag{7}$$

We now present our ScaledGD algorithm for high-order tensor completion in Algorithm 2.

Theoretical guarantees. The following theorem states that ScaledGD converges linearly at a constant rate as long as the sample size is sufficiently large. In the following, we denote $n_{(1)} \coloneqq \max\{n_1, n_2\}$ and $n_{(2)} \coloneqq \min\{n_1, n_2\}$.

Theorem 2. *Suppose that $\boldsymbol{\mathcal{X}}_\star$ is μ-incoherent, and that p satisfies $p \ge C(\frac{1}{\ell} \vee \mu r \kappa^4 \ell)\mu r \log(n_{(1)}\ell)/n_{(2)}$ for some sufficiently large constant C. Set the projection radius as $B = C_B\sqrt{\frac{\mu r}{\ell}}\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star)$ for some constant $C_B \ge 1.02$. If the step size obeys $0 < \eta \le 2/3$, then with high probability, for all $t \ge 0$, the iterates of ScaledGD in* (9) *satisfy*

$$\operatorname{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \le (1 - 0.6\eta)^t 0.02\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) \quad \text{and} \quad \|\boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t^T - \boldsymbol{\mathcal{X}}_\star\|_F \le (1 - 0.6\eta)^t 0.03\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).$$

Theorem 2 establishes that the distance $\operatorname{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star)$ contracts linearly at a constant rate, as long as the sample probability satisfies $p \gtrsim (\frac{1}{\ell} \vee \mu r \kappa^4 \ell)\mu r \log(n_{(1)}\ell)/n_{(2)}$. To reach ϵ-accuracy, i.e., $\|\boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t^T - \boldsymbol{\mathcal{X}}_\star\|_F \le \epsilon\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$, ScaledGD takes at most $\mathcal{O}(\log(1/\epsilon))$ iterations, which is independent of κ.

Algorithm 2 ScaledGD for order-d tensor completion with spectral initialization

Input: Partially observed data tensor $\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star)$, the corresponding matrices $\{\boldsymbol{U}_{n_k}\}_{k=3}^d$ of invertible transform $\mathfrak{L}$, the t-SVD rank r, learning rate η, and maximum number of iterations T.

Spectral initialization: Let $\boldsymbol{\mathcal{U}}_0 *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_0 *_{\mathfrak{L}} \boldsymbol{\mathcal{V}}_0^T$ be the top-r t-SVD of $\frac{1}{p}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star)$, and set

$$\begin{bmatrix} \boldsymbol{\mathcal{L}}_0 \\ \boldsymbol{\mathcal{R}}_0 \end{bmatrix} = \mathcal{P}_B\Big(\begin{bmatrix} \boldsymbol{\mathcal{U}}_0 *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_0^{\frac{1}{2}} \\ \boldsymbol{\mathcal{V}}_0 *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_0^{\frac{1}{2}} \end{bmatrix} \Big). \tag{8}$$

Scaled gradient updates: for $t = 0, 1, \ldots, T-1$ **do**

$$\begin{bmatrix} \boldsymbol{\mathcal{L}}_{t+1} \\ \boldsymbol{\mathcal{R}}_{t+1} \end{bmatrix} = \mathcal{P}_B\Big(\begin{bmatrix} \boldsymbol{\mathcal{L}}_t - \frac{\eta}{p}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t^T - \boldsymbol{\mathcal{X}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_t^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t)^{-1} \\ \boldsymbol{\mathcal{R}}_t - \frac{\eta}{p}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t^T - \boldsymbol{\mathcal{X}}_\star)^T *_{\mathfrak{L}} \boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} (\boldsymbol{\mathcal{L}}_t^T *_{\mathfrak{L}} \boldsymbol{\mathcal{L}}_t)^{-1} \end{bmatrix} \Big). \tag{9}$$

Output: The recovered low-rank tensor $\boldsymbol{\mathcal{X}}_T = \boldsymbol{\mathcal{L}}_T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_T^T$.

4.3 Proof Outline

In this section, we sketch the proof of our main theorem. First of all, to track the progress of ScaledGD throughout the entire trajectory, we use the same distance metric as in Wu (2025) to resolve the ambiguity in the t-SVD decomposition.

Definition 14 (Distance metric)**.** Let $\boldsymbol{\mathcal{F}} = \begin{bmatrix} \boldsymbol{\mathcal{L}} \\ \boldsymbol{\mathcal{R}} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times r\times n_3\times\cdots\times n_d}$ and $\boldsymbol{\mathcal{F}}_\star = \begin{bmatrix} \boldsymbol{\mathcal{L}}_\star \\ \boldsymbol{\mathcal{R}}_\star \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times r\times n_3\times\cdots\times n_d}$, denote

$$\mathrm{dist}(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}_\star) = \sqrt{\inf_{\boldsymbol{\mathcal{Q}}\in \mathrm{GL}(r)} \|(\boldsymbol{\mathcal{L}} *_{\mathfrak{L}} \boldsymbol{\mathcal{Q}} - \boldsymbol{\mathcal{L}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 + \|(\boldsymbol{\mathcal{R}} *_{\mathfrak{L}} \boldsymbol{\mathcal{Q}}^{-T} - \boldsymbol{\mathcal{R}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2}. \tag{10}$$

If the infimum is attained at the argument $\boldsymbol{\mathcal{Q}}$, it is called the optimal alignment tensor between $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{F}}_\star$.

We start with the following lemma that ensures the scaled projection in (7) satisfies both non-expansiveness and incoherence under the scaled metric.

Lemma 3. *Suppose that $\boldsymbol{\mathcal{X}}_\star$ is μ-incoherent, and $\mathrm{dist}(\widetilde{\boldsymbol{\mathcal{F}}}, \boldsymbol{\mathcal{F}}_\star) \le \frac{\epsilon}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$ for some $\epsilon < 1$. Set $B \ge (1+\epsilon)\sqrt{\frac{\mu r}{\ell}}\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star)$, then $\mathcal{P}_B(\widetilde{\boldsymbol{\mathcal{F}}})$ satisfies the non-expansiveness*

$$\mathrm{dist}(\mathcal{P}_B(\widetilde{\boldsymbol{\mathcal{F}}}), \boldsymbol{\mathcal{F}}_\star) \le \mathrm{dist}(\widetilde{\boldsymbol{\mathcal{F}}}, \boldsymbol{\mathcal{F}}_\star),$$

and the incoherence condition

$$\sqrt{n_1}\|\boldsymbol{\mathcal{L}} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}^T\|_{2,\infty} \vee \sqrt{n_2}\|\boldsymbol{\mathcal{R}} *_{\mathfrak{L}} \boldsymbol{\mathcal{L}}^T\|_{2,\infty} \le B.$$

Our next lemma states that as long as the sample complexity is large enough and the parameter B is set properly, the local convergence of Algorithm 2 is guaranteed.

Lemma 4. *Suppose that $\boldsymbol{\mathcal{X}}_\star$ is μ-incoherent, and $p \ge C(\frac{1}{\ell} \vee \mu r\kappa^4\ell)\mu r \log(n_{(1)}\ell)/n_{(2)}$ for some sufficiently large constant C. Set the projection radius as $B = C_B\sqrt{\frac{\mu r}{\ell}}\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star)$ for some constant $C_B \ge 1.02$. Under an event G which happens with high probability (i.e., at least $1 - c_1(n_1 \vee n_2)^{c_2}$), if the t-th iterate satisfies $\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \le \frac{0.02}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$, and the incoherence condition*

$$\sqrt{n_1}\|\boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t^T\|_{2,\infty} \vee \sqrt{n_2}\|\boldsymbol{\mathcal{R}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{L}}_t^T\|_{2,\infty} \le B,$$

then $\|\boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t^T - \boldsymbol{\mathcal{X}}_\star\|_F \le 1.5\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star)$. In addition, if the step size obeys $0 < \eta \le 2/3$, then the $(t+1)$-th iterate $\boldsymbol{\mathcal{F}}_{t+1}$ of the ScaledGD method in (9) of Algorithm 2 satisfies

$$\mathrm{dist}(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) \le (1 - 0.6\eta)\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star),$$

and the incoherence condition

$$\sqrt{n_1}\|\boldsymbol{\mathcal{L}}_{t+1} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_{t+1}^T\|_{2,\infty} \vee \sqrt{n_2}\|\boldsymbol{\mathcal{R}}_{t+1} *_{\mathfrak{L}} \boldsymbol{\mathcal{L}}_{t+1}^T\|_{2,\infty} \leq B.$$

As long as we can find an initialization that is close to the ground truth and satisfies the incoherence condition, Lemma 4 ensures that the iterates of ScaledGD remain incoherent and converge linearly. Such an initialization can be ensured via the spectral method, as stated below.

Lemma 5. *Suppose that $\boldsymbol{\mathcal{X}}_\star$ is μ-incoherent, then with high probability, the spectral initialization before projection $\widetilde{\boldsymbol{\mathcal{F}}}_0 := \begin{bmatrix} \boldsymbol{\mathcal{U}}_0 *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_0^{\frac{1}{2}} \\ \boldsymbol{\mathcal{V}}_0 *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_0^{\frac{1}{2}} \end{bmatrix}$ in (8) satisfies*

$$\operatorname{dist}(\widetilde{\boldsymbol{\mathcal{F}}}_0, \boldsymbol{\mathcal{F}}_\star) \leq c\Big(\frac{\mu r \log(n_{(1)}\ell)}{p\sqrt{n_1 n_2 \ell}} + \sqrt{\frac{\mu r \log(n_{(1)}\ell)}{n_{(2)} p \ell}}\Big) 5\sqrt{\frac{s_r}{\ell}}\kappa \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).$$

Therefore, as long as $p \geq C\mu r s_r \kappa^2 \log(n_{(1)}\ell)/(n_{(2)}\ell^2)$ for some sufficiently large constant C, the initial distance satisfies $\operatorname{dist}(\widetilde{\boldsymbol{\mathcal{F}}}_0, \boldsymbol{\mathcal{F}}_\star) \leq \frac{0.02}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$. One can then invoke Lemma 3 to see that $\boldsymbol{\mathcal{F}}_0 = \mathcal{P}_B(\widetilde{\boldsymbol{\mathcal{F}}}_0)$ meets the conditions required in Lemma 4 due to the non-expansiveness and incoherence properties of the projection operator. The proofs of the three supporting lemmas can be found in the Appendix.

5 Numerical Experiments

In this section, we present several experimental results on both synthetic and real data.

5.1 Synthetic Data Experiments

First, we verify the theoretical guarantee of order-d tensor completion in Theorem 2 through numerical simulations. In this experiment, we adopt Discrete Fourier Transform (DFT) and Discrete Cosine Transform (DCT) as the two invertible linear transforms. In order to generate the ground truth tensor $\boldsymbol{\mathcal{X}}_\star \in \mathbb{R}^{n\times n\times n_3\times\cdots\times n_d}$ with $\operatorname{rank}_{\text{t-SVD}}(\boldsymbol{\mathcal{X}}_\star) = r$, we first generate an $n \times r \times n_3 \times \cdots \times n_d$ tensor with i.i.d. random signs, and take its r left singular tensors as $\boldsymbol{\mathcal{U}}_\star$, and similarly for $\boldsymbol{\mathcal{V}}_\star$. We then set the diagonal entries in each frontal slice of the f-diagonal tensor $\boldsymbol{\mathcal{S}}_{\star,\mathfrak{L}} \in \mathbb{R}^{r\times r\times n_3\times\cdots\times n_d}$ to be linearly distributed from 1 to $1/\kappa$. In this way, the low-rank tensor generated by $\boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{U}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{V}}_\star^T$ will have the specified condition number κ and t-SVD rank r. In our experiment, we set $d = 4$, $n = 100$, $n_3 = n_4 = 50$ and $r = 10$. Then we randomly sample $pn^2n_3n_4$ elements with probability $p = 0.4$ from $\boldsymbol{\mathcal{X}}_\star$ to construct the known observations. The observation is $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star + \boldsymbol{\mathcal{W}})$, where $\boldsymbol{\mathcal{W}}_{i_1,i_2,i_3,i_4} \sim \mathcal{N}(0, \sigma_w^2)$ composed of i.i.d. Gaussian entries.

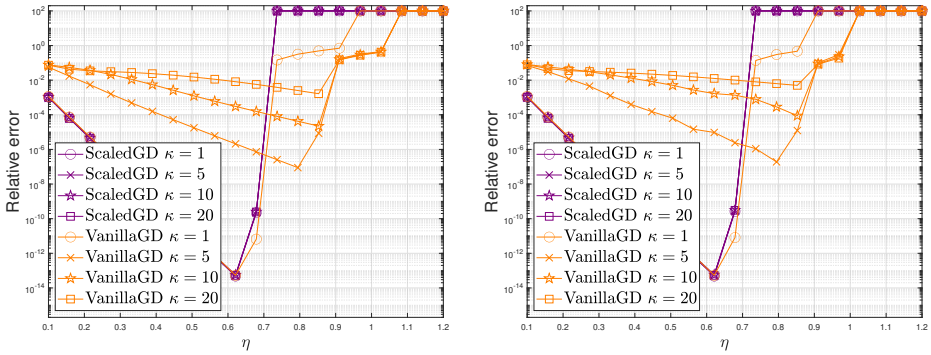
We compare the iteration complexity of ScaledGD against vanilla gradient descent (GD) with the same spectral initialization, and we use the following update rule of vanilla GD as

$$\boldsymbol{\mathcal{L}}_{t+1} = \boldsymbol{\mathcal{L}}_t - \eta_{\text{GD}}\nabla_{\boldsymbol{\mathcal{L}}} f(\boldsymbol{\mathcal{L}}_t, \boldsymbol{\mathcal{R}}_t) \quad \text{and} \quad \boldsymbol{\mathcal{R}}_{t+1} = \boldsymbol{\mathcal{R}}_t - \eta_{\text{GD}}\nabla_{\boldsymbol{\mathcal{R}}} f(\boldsymbol{\mathcal{L}}_t, \boldsymbol{\mathcal{R}}_t),$$

where $\eta_{\text{GD}} = \eta/\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star)$. We perform the scaled gradient updates without projections.

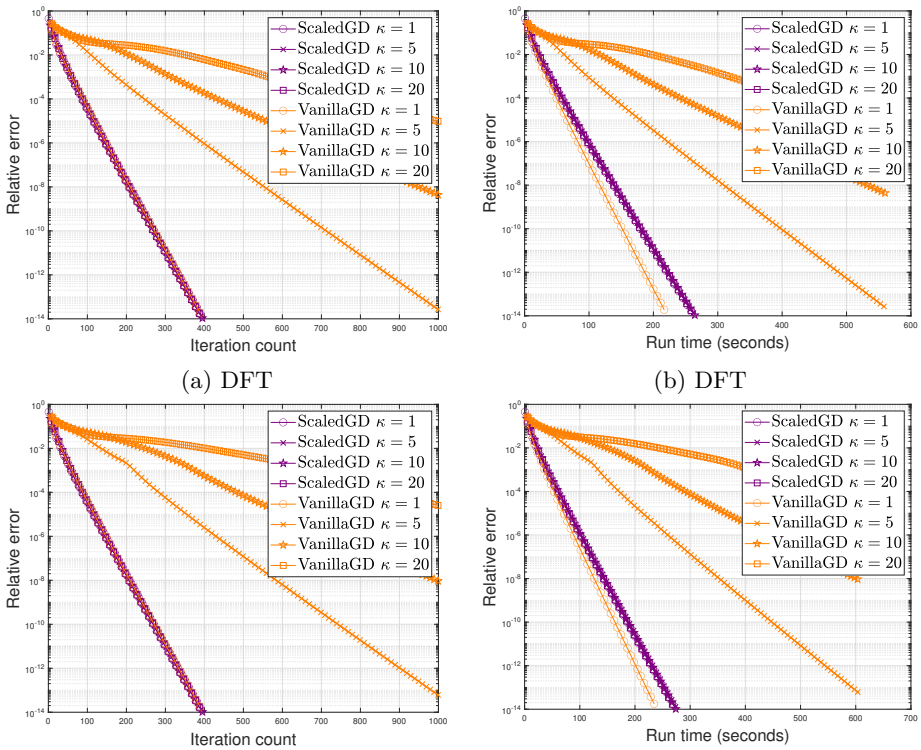
Figure 1 depicts the speeds of convergence for ScaledGD and vanilla GD under different step sizes in the noiseless setting, i.e., $\boldsymbol{\mathcal{W}} = \mathbf{0}$, where we run both algorithms for at most 300 iterations (the algorithm is terminated if the relative error exceeds 10^2) and plot the relative reconstruction error $\|\boldsymbol{\mathcal{X}}_T - \boldsymbol{\mathcal{X}}_\star\|_F/\|\boldsymbol{\mathcal{X}}_\star\|_F$ after these 300 iterations. It can be seen that even when the step size of vanilla GD is tuned to achieve its best performance, ScaledGD still performs much better than vanilla GD. Hence, we will fix $\eta = 0.5$ for the rest of the comparisons between ScaledGD and vanilla GD.

Figure 2 shows the relative reconstruction error $\|\boldsymbol{\mathcal{X}}_t - \boldsymbol{\mathcal{X}}_\star\|_F/\|\boldsymbol{\mathcal{X}}_\star\|_F$ of the two algorithms with respect to the iteration count and running time (in seconds) under different condition numbers $\kappa = 1, 5, 10, 20$ for the two transforms. This experiment verifies our theoretical finding that ScaledGD converges rapidly at a rate independent of κ, and this rate is the same as the one for vanilla GD under perfect conditioning $\kappa = 1$. Given



(a) DFT (b) DCT

Figure 1: The relative errors of ScaledGD and vanilla GD after 300 iterations with respect to different step sizes η from 0.1 to 1.2 under different condition numbers $\kappa = 1, 5, 10, 20$ with $n = 100$, $r = 10$, and $p = 0.4$.



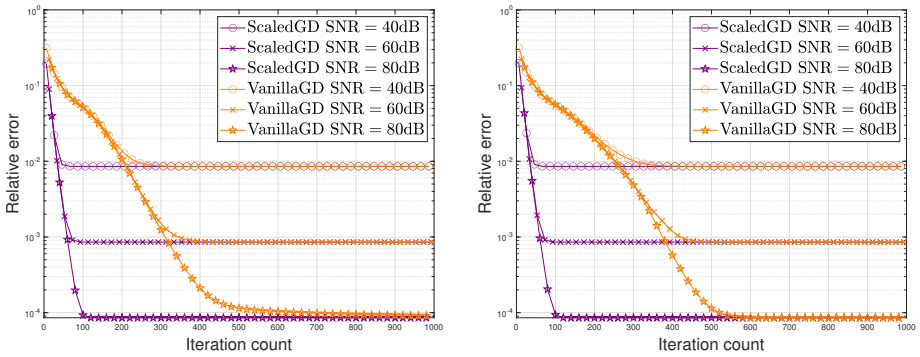
(a) DFT (b) DFT

(c) DCT (d) DCT

Figure 2: The relative errors of ScaledGD and vanilla GD with respect to (a/c) the iteration count and (b/d) run time (in seconds) under different condition numbers $\kappa = 1, 5, 10, 20$ with $n = 100$, $r = 10$, and $p = 0.4$.

that ScaledGD adds little overhead to the gradient computation, ScaledGD runs slightly slower than vanilla GD $\kappa = 1$. However, the convergence rate of vanilla GD collapses quickly when κ is even at a moderate level. It then turns out that ScaledGD carries over to the run time when $\kappa > 1$.

Next, we move to show that ScaledGD is robust to small additive noise. We denote the signal-to-noise ratio as $\mathrm{SNR} := 10\log_{10}\frac{\|\boldsymbol{\mathcal{X}}_\star\|_F^2}{n^2 n_3 n_4 \sigma_w^2}$ in dB. We plot the relative error $\|\boldsymbol{\mathcal{X}}_t - \boldsymbol{\mathcal{X}}_\star\|_F / \|\boldsymbol{\mathcal{X}}_\star\|_F$ with respect to the iteration count t in Figure 3 under $\kappa = 10$ and various $\mathrm{SNR} = 40, 60, 80$dB. We can see that ScaledGD converges much faster than vanilla GD and its convergence speed is not influenced by the noise levels.



(a) DFT (b) DCT

Figure 3: The relative errors of ScaledGD and vanilla GD with respect to the iteration count under signal-to-noise ratios SNR $= 40, 60, 80$dB with $n = 100$, $\kappa = 10$, $r = 10$, and $p = 0.4$.

Table 1: The PSNR, FSIM values and running time (in seconds) for StarPlus dataset.

Methods	SR = 10%			SR = 20%			SR = 30%		
	PSNR	FSIM	Time	PSNR	FSIM	Time	PSNR	FSIM	Time
HTNN-DFT	44.91	0.936	1945.0	45.63	0.943	1915.9	46.42	0.950	1812.7
HTNN-DCT	44.86	0.936	1997.2	45.57	0.943	1939.6	46.36	0.950	1817.4
ScaledGD-DFT	44.17	0.933	400.9	43.93	0.938	401.2	44.29	0.943	401.1
ScaledGD-DCT	44.16	0.932	513.4	43.92	0.938	518.2	44.28	0.943	518.2

5.2 Real Data Experiments

In this subsection, we compare the performance of ScaledGD with a tensor completion model that minimizes the order-d TNN (HTNN) (Qin et al., 2022) using the StarPlus fMRI dataset [1], which was collected Carnegie Mellon University's Center for Cognitive Brain Imaging. We again use two different linear transforms for ScaledGD, i.e., DFT and DCT. The corresponding methods of ScaledGD are called ScaledGD-DFT and ScaledGD-DCT for short. The StarPlus fMRI dataset includes 6 subjects with 80 trials for every single subject. Each trial is composed of a series of fMRI scans over a period of 16 time intervals spaced out over 500 milliseconds, where the subject is either reading a sentence or viewing a picture. The fMRI scan taken at each time slot consists of 8 axial slices with dimensions of 64×64 pixels. We orient the tensor such that the the trials are indexed in the second dimension, resulting in $\boldsymbol{\mathcal{X}}_\star \in \mathbb{R}^{64\times 480\times 64\times 8\times 16}$. We then add white Gaussian noise $\boldsymbol{\mathcal{W}}$ with $\boldsymbol{\mathcal{W}}_{i_1,\ldots,i_5} \sim \mathcal{N}(0, \frac{\|\boldsymbol{\mathcal{X}}_\star\|_F^2}{10^4 n_1 \ldots n_5})$ to the tensor $\boldsymbol{\mathcal{X}}_\star$, which results in SNR of $10\log_{10}\frac{\|\boldsymbol{\mathcal{X}}_\star\|_F^2}{\|\boldsymbol{\mathcal{W}}\|_F^2} = 40$dB. The sampling ratio (SR) is set to be SR $= [10\%, 20\%, 30\%]$. The t-SVD rank r in ScaledGD is set to be 10. To evaluate the tensor completion performance between the approximated tensor $\hat{\boldsymbol{\mathcal{X}}}$ and the original one $\boldsymbol{\mathcal{X}}_\star$, we adopt two metrics, namely, the Peak Signal-to-Noise Ratio (PSNR[2]) and the feature similarity (FSIM) (Zhang et al., 2011) (We also use the structural similarity (SSIM) (Wang et al., 2004) as the evaluation metric, and it is 1 for all the methods). It can be inferred from Table 1 that ScaledGD consumes less than a quarter of the computational time of HTNN but can achieve comparable performance to HTNN, which suggests an appealing paradigm for large-scale datasets.

6 Conclusion

In this paper, we proposed a scaled gradient descent (ScaledGD) algorithm for high-order tensor completion based upon the high-order t-SVD framework. The proposed algorithm has provable exact recovery and linear convergence guarantees, leading to a highly scalable approach especially when the ground truth tensor is

[1] `https://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/`

[2] PSNR $= 10\log_{10}(n_1 \times \cdots \times n_d \|\boldsymbol{\mathcal{X}}_\star\|_\infty^2 / \|\boldsymbol{\mathcal{X}}_\star - \hat{\boldsymbol{\mathcal{X}}}\|_F^2)$.

ill-conditioned. Extensive experimental results on synthetic and real data demonstrated the superiority of our method.

References

Talal Ahmed, Haroon Raja, and Waheed U. Bajwa. Tensor regression using low-rank and sparse Tucker decompositions. *SIAM Journal on Mathematics of Data Science*, 2(4):944–966, 2020.

Jonas Ballani and Lars Grasedyck. Tree adaptive approximation in the hierarchical tensor format. *SIAM Journal on Scientific Computing*, 36(4):A1415–A1431, 2014.

Johann A. Bengua, Ho N. Phien, Hoang Duong Tuan, and Minh N. Do. Efficient tensor completion for color image and video recovery: Low-rank tensor train. *IEEE Transactions on Image Processing*, 26(5): 2466–2479, 2017.

Adel Bibi and Bernard Ghanem. High order tensor formulation for convolutional sparse coding. In *IEEE International Conference on Computer Vision*, pp. 1790–1798, 2017.

Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

Wei Chen, Xiao Gong, and Nan Song. Nonconvex robust low-rank tensor reconstruction via an empirical Bayes method. *IEEE Transactions on Signal Processing*, 67(22):5785–5797, 2019.

Yudong Chen. Incoherence-optimal matrix completion. *IEEE Transactions on Information Theory*, 61(5): 2909–2923, 2015.

Yudong Chen and Martin J. Wainwright. Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. *arXiv preprint*, 2015. URL `https://arxiv.org/abs/1509.03025`.

Donald Goldfarb and Zhiwei Qin. Robust low-rank tensor recovery: Models and algorithms. *SIAM Journal on Matrix Analysis and Applications*, 35(1):225–253, 2014.

Christopher J. Hillar and Lek-Heng Lim. Most tensor problems are NP-hard. *Journal of the ACM*, 60(6): 1–39, 2013.

Huyan Huang, Yipeng Liu, Zhen Long, and Ce Zhu. Robust low-rank tensor ring completion. *IEEE Transactions on Computational Imaging*, 6:1117–1126, 2020.

Qiang Jiang and Michael Ng. Robust low-tubal-rank tensor completion via convex optimization. In *International Joint Conference on Artificial Intelligence*, pp. 2649–2655, 2019.

Tai-Xiang Jiang, Michael K. Ng, Xi-Le Zhao, and Ting-Zhu Huang. Framelet representation of tensor nuclear norm for third-order tensor completion. *IEEE Transactions on Image Processing*, 29:7233–7244, 2020.

Henk A. L. Kiers. Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, 14(3):105–122, 2000.

Misha E. Kilmer, Karen Braman, Ning Hao, and Randy C. Hoover. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis and Applications*, 34(1):148–172, 2013.

Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3): 455–500, 2009.

Hao Kong, Xingyu Xie, and Zhouchen Lin. t-schatten-p norm for low-rank tensor recovery. *IEEE Journal of Selected Topics in Signal Processing*, 12(6):1405–1419, 2018.

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.

Xiao Peng Li, Zhi-Yong Wang, Zhang-Lei Shi, Hing Cheung So, and Nicholas D. Sidiropoulos. Robust tensor completion via capped Frobenius norm. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7):9700–9712, 2024.

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, 2013.

Xiao-Yang Liu, Shuchin Aeron, Vaneet Aggarwal, and Xiaodong Wang. Low-tubal-rank tensor completion using alternating minimization. *IEEE Transactions on Information Theory*, 66(3):1714–1737, 2020.

Canyi Lu. Transforms based tensor robust PCA: Corrupted low-rank tensors recovery via convex optimization. In *IEEE/CVF International Conference on Computer Vision*, pp. 1145–1152, 2021.

Canyi Lu, Jiashi Feng, Zhouchen Lin, and Shuicheng Yan. Exact low tubal rank tensor recovery from Gaussian measurements. In *International Joint Conference on Artificial Intelligence*, pp. 2504–2510, 2018.

Canyi Lu, Xi Peng, and Yunchao Wei. Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5996–6004, 2019.

Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4):925–938, 2020.

Bamdev Mishra, K. Adithya Apuroop, and Rodolphe Sepulchre. A Riemannian geometry for low-rank matrix completion. *arXiv preprint*, 2012. URL `https://arxiv.org/abs/1211.1550`.

I. V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

Evangelos E. Papalexakis, Christos Faloutsos, and Nicholas D. Sidiropoulos. Tensors for data mining and data fusion: Models, applications, and scalable algorithms. *ACM Transactions on Intelligent Systems and Technology*, 8(2):1–44, 2016.

Wenjin Qin, Hailin Wang, Feng Zhang, Jianjun Wang, Xin Luo, and Tingwen Huang. Low-rank high-order tensor completion with applications in visual data. *IEEE Transactions on Image Processing*, 31:2433–2448, 2022.

Yuning Qiu, Guoxu Zhou, Qibin Zhao, and Shengli Xie. Noisy tensor completion via low-rank tensor ring. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):1127–1141, 2024.

Bernardino Romera-Paredes and Massimiliano Pontil. A new convex relaxation for tensor completion. In *Advances in Neural Information Processing Systems*, pp. 2967–2975, 2013.

Nicholas D. Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E. Papalexakis, and Christos Faloutsos. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, 2017.

Guangjing Song, Michael K. Ng, and Xiongjun Zhang. Robust tensor completion using transformed tensor singular value decomposition. *Numerical Linear Algebra with Applications*, 27(3):e2299, 2020.

Jared Tanner and Ke Wei. Low rank matrix completion by alternating steepest descent methods. *Applied and Computational Harmonic Analysis*, 40(2):417–429, 2016.

Tian Tong, Cong Ma, and Yuejie Chi. Accelerating ill-conditioned low-rank matrix estimation via scaled gradient descent. *Journal of Machine Learning Research*, 22(150):1–63, 2021.

Tian Tong, Cong Ma, Ashley Prater-Bennette, Erin Tripp, and Yuejie Chi. Scaling and scalability: Provable nonconvex low-rank tensor estimation from incomplete measurements. *Journal of Machine Learning Research*, 23(163):1–77, 2022.

Ledyard R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

Hailin Wang, Feng Zhang, Jianjun Wang, Tingwen Huang, Jianwen Huang, and Xinling Liu. Generalized nonconvex approach for low-tubal-rank tensor recovery. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8):3305–3319, 2022.

Hailin Wang, Jiangjun Peng, Wenjin Qin, Jianjun Wang, and Deyu Meng. Guaranteed tensor recovery fused low-rankness and smoothness. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9): 10990–11007, 2023.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

Tong Wu. Guaranteed nonconvex low-rank tensor estimation via scaled gradient descent. *arXiv preprint*, 2025. URL `https://www.arxiv.org/abs/2501.01696`.

Longhao Yuan, Qibin Zhao, Lihua Gui, and Jianting Cao. High-order tensor completion via gradient-based optimization under tensor train format. *Signal Processing: Image Communication*, 73:53–61, 2019.

Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011.

Xiaoqin Zhang, Di Wang, Zhengyuan Zhou, and Yi Ma. Robust low-rank tensor recovery with rectification and alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):238–255, 2021.

Xiongjun Zhang and Michael K. Ng. Low rank tensor completion with Poisson observations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4239–4251, 2022.

Zemin Zhang and Shuchin Aeron. Exact tensor completion using t-SVD. *IEEE Transactions on Signal Processing*, 65(6):1511–1526, 2017.

Zemin Zhang, Gregory Ely, Shuchin Aeron, Ning Hao, and Misha Kilmer. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3842–3849, 2014.

Qibin Zhao, Guoxu Zhou, Shengli Xie, Liqing Zhang, and Andrzej Cichocki. Tensor ring decomposition. *arXiv preprint*, 2016. URL `https://arxiv.org/abs/1606.05535`.

Qinqing Zheng and John Lafferty. Convergence analysis for rectangular matrix completion using Burer-Monteiro factorization and gradient descent. *arXiv preprint*, 2016. URL `https://arxiv.org/abs/1605.07051`.

Pan Zhou, Canyi Lu, Zhouchen Lin, and Chao Zhang. Tensor factorization for low-rank tensor completion. *IEEE Transactions on Image Processing*, 27(3):1152–1163, 2018.

A Proof of Main Result

In this section, we provide detailed proofs of the main result in Theorem 2. We first give some definitions and properties which will be used in the proofs.

A.1 Technical Lemmas

This section gathers several technical lemmas that will be used in the proofs. We use bold calligraphic letters with arrows on top to denote tensor columns of size $n_1 \times 1 \times n_3 \times \cdots \times n_d$, e.g., $\vec{\boldsymbol{\mathcal{A}}}$. We define the $\ell_{\infty,2}$-norm and $\ell_{2\diamond\infty}$-norm of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ as

$$\|\boldsymbol{\mathcal{A}}\|_{\infty,2} = \max\{\max_{i_1} \|\boldsymbol{\mathcal{A}}(i_1, :, \ldots, :)\|_F, \max_{i_2} \|\boldsymbol{\mathcal{A}}(:, i_2, \ldots, :)\|_F\},$$

and $\|\boldsymbol{\mathcal{A}}\|_{2\diamond\infty} = \max_{i_1,i_3,\ldots,i_d} \|\boldsymbol{\mathcal{A}}(i_1,:,i_3,\ldots,i_d)\|_F$, respectively. The spectral norm of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ is defined as $\|\boldsymbol{\mathcal{A}}\| = \|\texttt{bdiag}(\boldsymbol{\mathcal{A}}_{\mathfrak{L}})\|$.

A.1.1 Tensor Algebra

We begin with introducing some relevant algebraic properties of t-SVD. Based on the assumption (1), we have the following properties:

$$\|\boldsymbol{\mathcal{A}}\|_F = \frac{1}{\sqrt{\ell}}\|\texttt{bdiag}(\boldsymbol{\mathcal{A}}_{\mathfrak{L}})\|_F \quad \text{and} \quad \langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\rangle = \frac{1}{\ell}\langle \texttt{bdiag}(\boldsymbol{\mathcal{A}}_{\mathfrak{L}}), \texttt{bdiag}(\boldsymbol{\mathcal{B}}_{\mathfrak{L}})\rangle. \tag{11}$$

Definition 15. Let $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ with $\text{rank}_{\text{t-SVD}}(\boldsymbol{\mathcal{M}}) = r$ and its skinny t-SVD be $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{U}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}} *_{\mathfrak{L}} \boldsymbol{\mathcal{V}}^T$. Define $\boldsymbol{T}$ by the set

$$\boldsymbol{T} = \{\boldsymbol{\mathcal{U}} *_{\mathfrak{L}} \boldsymbol{\mathcal{Z}}^T + \boldsymbol{\mathcal{W}} *_{\mathfrak{L}} \boldsymbol{\mathcal{V}}^T | \boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_2\times r\times n_3\times\cdots\times n_d}, \boldsymbol{\mathcal{W}} \in \mathbb{R}^{n_1\times r\times n_3\times\cdots\times n_d}\}, \tag{12}$$

and by $\boldsymbol{T}^{\perp}$ its orthogonal complement.

Lemma 6. *Let $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3\times\cdots\times n_d}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{n_4\times n_2\times n_3\times\cdots\times n_d}$ be two tensors, then*

$$\|\boldsymbol{\mathcal{A}} *_{\mathfrak{L}} \boldsymbol{\mathcal{B}}^T\|_\infty \le \sqrt{\ell}\|\boldsymbol{\mathcal{A}}\|_{2,\infty}\|\boldsymbol{\mathcal{B}}\|_{2,\infty}.$$

Proof. By the definition of the high-order t-product, we have

$$\begin{aligned}
\|\boldsymbol{\mathcal{A}} *_{\mathfrak{L}} \boldsymbol{\mathcal{B}}^T\|_\infty &= \max_{i_1,i_2} \|\sum_{j=1}^{n_2} \boldsymbol{\mathcal{A}}(i_1,j,:,\ldots,:) *_{\mathfrak{L}} \boldsymbol{\mathcal{B}}(i_2,j,:,\ldots,:)^T\|_\infty \\
&\le \max_{i_1,i_2} \|\sum_{j=1}^{n_2} \boldsymbol{\mathcal{A}}(i_1,j,:,\ldots,:) *_{\mathfrak{L}} \boldsymbol{\mathcal{B}}(i_2,j,:,\ldots,:)^T\|_F \\
&\le \max_{i_1,i_2} \sum_{j=1}^{n_2} \|\boldsymbol{\mathcal{A}}(i_1,j,:,\ldots,:) *_{\mathfrak{L}} \boldsymbol{\mathcal{B}}(i_2,j,:,\ldots,:)^T\|_F \\
&= \max_{i_1,i_2} \sum_{j=1}^{n_2} \frac{1}{\sqrt{\ell}}\|\texttt{bdiag}(\boldsymbol{\mathcal{A}}_{\mathfrak{L}}(i_1,j,:,\ldots,:)) \cdot \texttt{bdiag}(\boldsymbol{\mathcal{B}}_{\mathfrak{L}}(i_2,j,:,\ldots,:)^T)\|_F.
\end{aligned}$$

Notice that each frontal slice of $\boldsymbol{\mathcal{A}}_{\mathfrak{L}}(i_1,j,:,\ldots,:)$ and $\boldsymbol{\mathcal{B}}_{\mathfrak{L}}(i_2,j,:,\ldots,:)$ are just scalars, thus both $\texttt{bdiag}(\boldsymbol{\mathcal{A}}_{\mathfrak{L}}(i_1,j,:,\ldots,:)) \in \mathbb{C}^{n_3\ldots n_d\times n_3\ldots n_d}$ and $\texttt{bdiag}(\boldsymbol{\mathcal{B}}_{\mathfrak{L}}(i_2,j,:,\ldots,:)^T) \in \mathbb{C}^{n_3\ldots n_d\times n_3\ldots n_d}$ are diagonal matrices. Hence

$$\begin{aligned}
\|\boldsymbol{\mathcal{A}} *_{\mathfrak{L}} \boldsymbol{\mathcal{B}}^T\|_\infty &\le \max_{i_1,i_2} \sum_{j=1}^{n_2} \frac{1}{\sqrt{\ell}}\|\texttt{bdiag}(\boldsymbol{\mathcal{A}}_{\mathfrak{L}}(i_1,j,:,\ldots,:))\|_F\|\texttt{bdiag}(\boldsymbol{\mathcal{B}}_{\mathfrak{L}}(i_2,j,:,\ldots,:)^T)\|_F \\
&= \max_{i_1,i_2} \sum_{j=1}^{n_2} \sqrt{\ell}\|\texttt{bdiag}(\boldsymbol{\mathcal{A}}(i_1,j,:,\ldots,:))\|_F\|\texttt{bdiag}(\boldsymbol{\mathcal{B}}(i_2,j,:,\ldots,:))\|_F \\
&\le \max_{i_1,i_2} \sqrt{\ell}\|\boldsymbol{\mathcal{A}}(i_1,:,\ldots,:)\|_F\|\boldsymbol{\mathcal{B}}(i_2,:,\ldots,:)\|_F \\
&\le \sqrt{\ell}\|\boldsymbol{\mathcal{A}}\|_{2,\infty}\|\boldsymbol{\mathcal{B}}\|_{2,\infty}.
\end{aligned}$$

□

Lemma 7. *Let $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3\times\cdots\times n_d}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{n_4\times n_2\times n_3\times\cdots\times n_d}$ be two tensors. Assume the multi-rank of $\boldsymbol{\mathcal{B}}$ is $\boldsymbol{r}$ and let $s_r = \sum_{k=1}^{n_3\cdots n_d} r_k$, then*

$$\|\boldsymbol{\mathcal{A}} *_{\mathfrak{L}} \boldsymbol{\mathcal{B}}\|_F \ge \|\boldsymbol{\mathcal{A}}\|_F\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{B}}) \quad \text{and} \quad \|\boldsymbol{\mathcal{A}} *_{\mathfrak{L}} \boldsymbol{\mathcal{B}}\|_F \le \|\boldsymbol{\mathcal{A}}\|_F\|\boldsymbol{\mathcal{B}}\|.$$

Moreover,

$$\|\boldsymbol{\mathcal{A}} *_{\mathfrak{L}} \boldsymbol{\mathcal{B}}\|_{2,\infty} \ge \|\boldsymbol{\mathcal{A}}\|_{2,\infty}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{B}}) \quad \text{and} \quad \|\boldsymbol{\mathcal{A}} *_{\mathfrak{L}} \boldsymbol{\mathcal{B}}\|_{2,\infty} \le \|\boldsymbol{\mathcal{A}}\|_{2,\infty}\|\boldsymbol{\mathcal{B}}\|.$$

Proof. The proof is identical to that for Lemma 10 of Wu (2025). □

A.1.2 Distance Metric

Lemma 8. *Fix any factor tensor* $\boldsymbol{\mathcal{F}} = \begin{bmatrix} \boldsymbol{\mathcal{L}} \\ \boldsymbol{\mathcal{R}} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times r\times n_3\times\cdots\times n_d}$. *Suppose that*

$$\operatorname{dist}(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}_\star) < \frac{1}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star), \tag{13}$$

then the optimal alignment tensor $\boldsymbol{\mathcal{Q}}$ *between* $\boldsymbol{\mathcal{F}}$ *and* $\boldsymbol{\mathcal{F}}_\star$ *exists.*

Proof. Given the condition (13), one knows that there must exist a tensor $\widetilde{\boldsymbol{\mathcal{Q}}} \in \mathrm{GL}(r)$ such that

$$\begin{aligned}
&\left\|\left(\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{L}}))\cdot\mathtt{bdiag}(\mathfrak{L}(\widetilde{\boldsymbol{\mathcal{Q}}}))-\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{L}}_\star))\right)\cdot\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}))\cdot\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{S}}_\star))\right\|_F^2 \\
&\quad+\left\|\left(\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{R}}))\cdot\mathtt{bdiag}(\mathfrak{L}(\widetilde{\boldsymbol{\mathcal{Q}}}^{-T}))-\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{R}}_\star))\right)\cdot\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}))\cdot\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{S}}_\star))\right\|_F^2 \\
&=\ell(\|(\boldsymbol{\mathcal{L}}*_{\mathfrak{L}}\widetilde{\boldsymbol{\mathcal{Q}}}-\boldsymbol{\mathcal{L}}_\star)*_{\mathfrak{L}}\boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2+\|(\boldsymbol{\mathcal{R}}*_{\mathfrak{L}}\widetilde{\boldsymbol{\mathcal{Q}}}^{-T}-\boldsymbol{\mathcal{R}}_\star)*_{\mathfrak{L}}\boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2) \\
&\le\epsilon^2\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star)
\end{aligned}$$

for some ϵ obeying $0<\epsilon<1$. In light of the relation $\|\boldsymbol{AB}\|_F \ge \|\boldsymbol{A}\|_F\sigma_{\min}(\boldsymbol{B})$,

$$\begin{aligned}
&\left\|\left(\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{L}}))\cdot\mathtt{bdiag}(\mathfrak{L}(\widetilde{\boldsymbol{\mathcal{Q}}}))-\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{L}}_\star))\right)\cdot\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}))\right\|_F^2 \\
&\quad+\left\|\left(\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{R}}))\cdot\mathtt{bdiag}(\mathfrak{L}(\widetilde{\boldsymbol{\mathcal{Q}}}^{-T}))-\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{R}}_\star))\right)\cdot\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}))\right\|_F^2\le\epsilon^2.
\end{aligned}$$

It further implies that

$$\begin{aligned}
&\left\|\left(\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{L}}))\cdot\mathtt{bdiag}(\mathfrak{L}(\widetilde{\boldsymbol{\mathcal{Q}}}))-\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{L}}_\star))\right)\cdot\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}))\right\| \\
&\quad\vee\left\|\left(\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{R}}))\cdot\mathtt{bdiag}(\mathfrak{L}(\widetilde{\boldsymbol{\mathcal{Q}}}^{-T}))-\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{R}}_\star))\right)\cdot\mathtt{bdiag}(\mathfrak{L}(\boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}))\right\|\le\epsilon.
\end{aligned}$$

The rest of the proof is the same as the one in Tong et al. (2021, Lemma 22). □

Further, following the proof in Tong et al. (2021, Lemma 24), we connect the distance metric $\operatorname{dist}(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}_\star)$ to the Frobenius norm in Lemma 9.

Lemma 9. *For any factor tensor* $\boldsymbol{\mathcal{F}} = \begin{bmatrix} \boldsymbol{\mathcal{L}} \\ \boldsymbol{\mathcal{R}} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times r\times n_3\times\cdots\times n_d}$, *the distance between* $\boldsymbol{\mathcal{F}}$ *and* $\boldsymbol{\mathcal{F}}_\star$ *satisfies*

$$\operatorname{dist}(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}_\star) \le \left(\sqrt{2}+1\right)^{\frac{1}{2}}\|\boldsymbol{\mathcal{L}}*_{\mathfrak{L}}\boldsymbol{\mathcal{R}}^T-\boldsymbol{\mathcal{X}}_\star\|_F.$$

A.1.3 Tensor Perturbation Bounds

Following the proof in Lemma 19 of Wu (2025), we have the following bound.

Lemma 10. *For any* $\boldsymbol{\mathcal{L}} \in \mathbb{R}^{n_1\times r\times n_3\times\cdots\times n_d}$, $\boldsymbol{\mathcal{R}} \in \mathbb{R}^{n_2\times r\times n_3\times\cdots\times n_d}$, *denote* $\boldsymbol{\mathcal{L}}_\triangle \coloneqq \boldsymbol{\mathcal{L}}-\boldsymbol{\mathcal{L}}_\star$ *and* $\boldsymbol{\mathcal{R}}_\triangle \coloneqq \boldsymbol{\mathcal{R}}-\boldsymbol{\mathcal{R}}_\star$, *then*

$$\begin{aligned}
\|\boldsymbol{\mathcal{L}}*_{\mathfrak{L}}\boldsymbol{\mathcal{R}}^T-\boldsymbol{\mathcal{X}}_\star\|_F &\le \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}}\boldsymbol{\mathcal{R}}_\star^T\|_F+\|\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}}\boldsymbol{\mathcal{R}}_\triangle^T\|_F+\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}}\boldsymbol{\mathcal{R}}_\triangle^T\|_F \\
&\le\left(1+\frac{1}{2}(\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}}\boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|\vee\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}}\boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|)\right) \\
&\qquad\left(\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}}\boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F+\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}}\boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F\right).
\end{aligned}$$

A.2 Proof of Lemma 3

Lemma 11 (Tong et al. (2021), Claim 5)**.** *For tensor columns* $\overrightarrow{\boldsymbol{\mathcal{A}}}, \overrightarrow{\boldsymbol{\mathcal{A}}}_\star \in \mathbb{R}^{n_1 \times 1 \times n_3 \times \cdots \times n_d}$ *and* $\lambda \geq \|\overrightarrow{\boldsymbol{\mathcal{A}}}_\star\|_F / \|\overrightarrow{\boldsymbol{\mathcal{A}}}\|_F$*, it holds that*

$$\|(1 \wedge \lambda)\overrightarrow{\boldsymbol{\mathcal{A}}} - \overrightarrow{\boldsymbol{\mathcal{A}}}_\star\|_F \leq \|\overrightarrow{\boldsymbol{\mathcal{A}}} - \overrightarrow{\boldsymbol{\mathcal{A}}}_\star\|_F.$$

Denote the optimal alignment tensor between $\widetilde{\boldsymbol{\mathcal{F}}}$ and $\boldsymbol{\mathcal{F}}_\star$ as $\widetilde{\boldsymbol{\mathcal{Q}}}$, whose existence is guaranteed by Lemma 8. Let $\mathcal{P}_B(\widetilde{\boldsymbol{\mathcal{F}}}) = \begin{bmatrix} \widetilde{\boldsymbol{\mathcal{L}}} \\ \widetilde{\boldsymbol{\mathcal{R}}} \end{bmatrix}$, by the definition of $\operatorname{dist}(\mathcal{P}_B(\widetilde{\boldsymbol{\mathcal{F}}}), \boldsymbol{\mathcal{F}}_\star)$, we have

$$\begin{aligned}\operatorname{dist}^2(\mathcal{P}_B(\widetilde{\boldsymbol{\mathcal{F}}}), \boldsymbol{\mathcal{F}}_\star) \leq & \sum_{i=1}^{n_1} \|\boldsymbol{\mathcal{L}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} - (\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}})(i,:,\ldots,:)\|_F^2 \\ & + \sum_{j=1}^{n_2} \|\boldsymbol{\mathcal{R}}(j,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}}^{-T} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} - (\boldsymbol{\mathcal{R}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}})(j,:,\ldots,:)\|_F^2. \end{aligned} \tag{14}$$

Recall that the condition $\operatorname{dist}(\widetilde{\boldsymbol{\mathcal{F}}}, \boldsymbol{\mathcal{F}}_\star) \leq \frac{\epsilon}{\sqrt{\ell}} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$ implies

$$\|(\widetilde{\boldsymbol{\mathcal{L}}} *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}} - \boldsymbol{\mathcal{L}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\| \vee \|(\widetilde{\boldsymbol{\mathcal{R}}} *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}}^{-T} - \boldsymbol{\mathcal{R}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\| \leq \epsilon.$$

Combining this with $\boldsymbol{\mathcal{R}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}} = \boldsymbol{\mathcal{V}}_\star$, we arrive at

$$\begin{aligned}\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{R}}}^T\|_F &\leq \|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \|\widetilde{\boldsymbol{\mathcal{R}}} *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}}^{-T} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\| \\ &\leq \|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \Big(\|\boldsymbol{\mathcal{V}}_\star\| + \|(\widetilde{\boldsymbol{\mathcal{R}}} *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}}^{-T} - \boldsymbol{\mathcal{R}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|\Big) \\ &\leq (1+\epsilon)\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F.\end{aligned}$$

In addition, the μ-incoherence of $\boldsymbol{\mathcal{X}}_\star$ yields

$$\sqrt{n_1}\|(\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}})(i,:,\ldots,:)\|_F \leq \sqrt{n_1}\|\boldsymbol{\mathcal{U}}_\star\|_{2,\infty}\|\boldsymbol{\mathcal{S}}_\star\| \leq \sqrt{\frac{\mu r}{\ell}}\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star) \leq \frac{B}{1+\epsilon},$$

where the last inequality follows from the choice of B. Taking the above two relations to reach

$$\frac{B}{\sqrt{n_1}\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{R}}}^T\|_F} \geq \frac{\|(\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}})(i,:,\ldots,:)\|_F}{\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F}.$$

We can then apply Lemma 11 with $\overrightarrow{\boldsymbol{\mathcal{A}}} := \widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}$, $\overrightarrow{\boldsymbol{\mathcal{A}}}_\star := (\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}})(i,:,\ldots,:)$, and $\lambda := \frac{B}{\sqrt{n_1}\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{R}}}^T\|_F}$ to obtain

$$\begin{aligned}&\|\boldsymbol{\mathcal{L}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} - (\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}})(i,:,\ldots,:)\|_F^2 \\ =&\Bigg\|\Bigg(1 \wedge \frac{B}{\sqrt{n_1}\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{R}}}^T\|_F}\Bigg)\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} - (\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}})(i,:,\ldots,:)\Bigg\|_F^2 \\ \leq&\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} - (\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}})(i,:,\ldots,:)\|_F^2.\end{aligned}$$

Following a similar argument for $\boldsymbol{\mathcal{R}}$, we conclude that

$$\begin{aligned}\operatorname{dist}^2(\mathcal{P}_B(\widetilde{\boldsymbol{\mathcal{F}}}), \boldsymbol{\mathcal{F}}_\star) \leq & \sum_{i=1}^{n_1} \|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} - (\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}})(i,:,\ldots,:)\|_F^2 \\ & + \sum_{j=1}^{n_2} \|\widetilde{\boldsymbol{\mathcal{R}}}(j,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{Q}}}^{-T} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} - (\boldsymbol{\mathcal{R}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}})(j,:,\ldots,:)\|_F^2 \leq \operatorname{dist}^2(\widetilde{\boldsymbol{\mathcal{F}}}, \boldsymbol{\mathcal{F}}_\star).\end{aligned}$$

We move on to the incoherence condition. For any $i \in [n_1]$, one has

$$
\begin{aligned}
&\|\boldsymbol{\mathcal{L}}(i,:,\ldots,:) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}^T\|_F^2 \\
=&\sum_{j=1}^{n_2} \|\boldsymbol{\mathcal{L}}(i,:,\ldots,:) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}(j,:,\ldots,:)^T\|_F^2 \\
=&\sum_{j=1}^{n_2} \Big(1 \wedge \frac{B}{\sqrt{n_1}\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{R}}}^T\|_F}\Big)^2 \|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{R}}}(j,:,\ldots,:)^T\|_F^2 \Big(1 \wedge \frac{B}{\sqrt{n_2}\|\widetilde{\boldsymbol{\mathcal{R}}}(j,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{L}}}^T\|_F}\Big)^2 \\
\overset{\text{(i)}}{\leq}&\Big(1 \wedge \frac{B}{\sqrt{n_1}\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{R}}}^T\|_F}\Big)^2 \sum_{j=1}^{n_2} \|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{R}}}(j,:,\ldots,:)^T\|_F^2 \\
=&\Big(1 \wedge \frac{B}{\sqrt{n_1}\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{R}}}^T\|_F}\Big)^2 \|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:) *_{\mathfrak{L}} \widetilde{\boldsymbol{\mathcal{R}}}^T\|_F^2 \\
\overset{\text{(ii)}}{\leq}&\frac{B^2}{n_1},
\end{aligned}
$$

where (i) follows from $1 \wedge \frac{B}{\sqrt{n_2}\|\widetilde{\boldsymbol{\mathcal{R}}}(j,:,\ldots,:)*_{\mathfrak{L}}\widetilde{\boldsymbol{\mathcal{L}}}^T\|_F} \leq 1$, and (ii) follows from $1 \wedge \frac{B}{\sqrt{n_1}\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:)*_{\mathfrak{L}}\widetilde{\boldsymbol{\mathcal{R}}}^T\|_F} \leq \frac{B}{\sqrt{n_1}\|\widetilde{\boldsymbol{\mathcal{L}}}(i,:,\ldots,:)*_{\mathfrak{L}}\widetilde{\boldsymbol{\mathcal{R}}}^T\|_F}$. Similarly, one can also have $\|\boldsymbol{\mathcal{R}}(j,:,\ldots,:) *_{\mathfrak{L}} \boldsymbol{\mathcal{L}}^T\|_F^2 \leq \frac{B^2}{n_2}$. Combining these two bounds completes the proof.

A.3 Proof of Lemma 4

We gather several useful inequalities regarding the operator $\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\cdot)$ for the Bernoulli observation model.

Lemma 12 (Qin et al. (2022), Lemma VI.8)**.** *Suppose that $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is fixed, and $\boldsymbol{\Omega} \sim \mathrm{Ber}(p)$. Then with high probability,*

$$
\|(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{Z}})\| \leq c\Big(\frac{\log(n_{(1)}\ell)}{p}\|\boldsymbol{\mathcal{Z}}\|_\infty + \sqrt{\frac{\log(n_{(1)}\ell)}{p}}\|\boldsymbol{\mathcal{Z}}\|_{\infty,2}\Big),
$$

for some numerical constant $c > 0$.

Lemma 13 (Wang et al. (2023), Lemma 11)**.** *Suppose that $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is fixed, and $\boldsymbol{\Omega} \sim \mathrm{Ber}(p)$. Then with high probability,*

$$
\|(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{Z}})\| \leq c\sqrt{\frac{n_{(1)}\ell\log(n_{(1)}\ell)}{p}}\|\boldsymbol{\mathcal{Z}}\|_\infty,
$$

for some numerical constant $c > 0$.

Next, following the proof of Lemma 10 in Zheng & Lafferty (2016), we have the restricted strong convexity and smoothness of the observation operator for tensors in $\boldsymbol{T}$.

Lemma 14. *Suppose that $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in \boldsymbol{T}$ are fixed tensors and $\boldsymbol{\Omega} \sim \mathrm{Ber}(p)$. Then with high probability,*

$$
p(1-\epsilon)\|\boldsymbol{\mathcal{A}}\|_F^2 \leq \|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{A}})\|_F^2 \leq p(1+\epsilon)\|\boldsymbol{\mathcal{A}}\|_F^2. \tag{15}
$$

Consequently,

$$
|p^{-1}\langle \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{A}}), \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{B}})\rangle - \langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\rangle| \leq \epsilon\|\boldsymbol{\mathcal{A}}\|_F\|\boldsymbol{\mathcal{B}}\|_F, \tag{16}
$$

provided that $p \geq c\epsilon^{-2}\mu r\log(n_{(1)}\ell)/(n_{(2)}\ell)$ for some numerical constant $c > 0$.

We then have the following simple corollary.

Corollary 15. *Suppose that $\boldsymbol{\mathcal{X}}_\star$ is μ-incoherent, and $p \gtrsim \mu r \log(n_{(1)}\ell)/(n_{(2)}\ell)$. Then with high probability,*

$$|\langle (p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T + \boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T), \boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T + \boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T\rangle| \\ \leq c\sqrt{\frac{\mu r \log(n_{(1)}\ell)}{p n_{(2)}\ell}} \|\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T + \boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T\|_F \|\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T + \boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T\|_F,$$

simultaneously for all $\boldsymbol{\mathcal{L}}_A, \boldsymbol{\mathcal{L}}_B \in \mathbb{R}^{n_1 \times r \times n_3 \times \cdots \times n_d}$ and $\boldsymbol{\mathcal{R}}_A, \boldsymbol{\mathcal{R}}_B \in \mathbb{R}^{n_2 \times r \times n_3 \times \cdots \times n_d}$, where $c > 0$ is some numerical constant.

Lemma 16. *Suppose that $p \gtrsim \log(n_{(1)}\ell)/n_{(2)}$. Then with high probability,*

$$|\langle (p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T), \boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T\rangle| \\ \leq c\ell^{\frac{3}{2}} \sqrt{\frac{n_{(1)} \log(n_{(1)}\ell)}{p}} \\ \Big(\|\boldsymbol{\mathcal{L}}_A\|_{2\diamond\infty}\|\boldsymbol{\mathcal{L}}_B\|_F \wedge \|\boldsymbol{\mathcal{L}}_A\|_F\|\boldsymbol{\mathcal{L}}_B\|_{2\diamond\infty}\Big)\Big(\|\boldsymbol{\mathcal{R}}_A\|_{2\diamond\infty}\|\boldsymbol{\mathcal{R}}_B\|_F \wedge \|\boldsymbol{\mathcal{R}}_A\|_F\|\boldsymbol{\mathcal{R}}_B\|_{2\diamond\infty}\Big),$$

simultaneously for all $\boldsymbol{\mathcal{L}}_A, \boldsymbol{\mathcal{L}}_B \in \mathbb{R}^{n_1 \times r \times n_3 \times \cdots \times n_d}$ and $\boldsymbol{\mathcal{R}}_A, \boldsymbol{\mathcal{R}}_B \in \mathbb{R}^{n_2 \times r \times n_3 \times \cdots \times n_d}$, where $c > 0$ is some universal constant.

Proof. First, for any $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}$, we have

$$|\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\rangle| = \frac{1}{\ell}|\langle \texttt{bdiag}(\boldsymbol{\mathcal{A}}_{\mathfrak{L}}), \texttt{bdiag}(\boldsymbol{\mathcal{B}}_{\mathfrak{L}})\rangle| \leq \frac{1}{\ell}\|\texttt{bdiag}(\boldsymbol{\mathcal{A}}_{\mathfrak{L}})\|\|\texttt{bdiag}(\boldsymbol{\mathcal{B}}_{\mathfrak{L}})\|_* = \|\boldsymbol{\mathcal{A}}\|\|\boldsymbol{\mathcal{B}}\|_{\circledast,\mathfrak{L}}.$$

Hence,

$$\begin{aligned} &|p^{-1}\langle \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T), \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T)\rangle - \langle \boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T, \boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T\rangle| \\ =&|\langle (p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{J}}), \big((\boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T) \circ (\boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T)\big)\rangle| \\ \leq&\|(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{J}})\|\|(\boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T) \circ (\boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T)\|_{\circledast,\mathfrak{L}}, \end{aligned} \tag{17}$$

where $\boldsymbol{\mathcal{J}}$ denotes tensor with all-one entries and $\circ$ denotes the Hadamard (elementwise) product. Following Lemma 13, let $\epsilon = c\sqrt{\frac{n_{(1)}\ell \log(n_{(1)}\ell)}{p}}$, then with high probability, we have

$$\|(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{J}})\| \leq \epsilon,$$

provided that $p \geq c\log(n_{(1)}\ell)/n_{(2)}$. Given the definition of tensor nuclear norm, we have

$$\begin{aligned} &\|(\boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T) \circ (\boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T)\|_{\circledast,\mathfrak{L}} \\ &= \frac{1}{\ell}\sum_{i_3=1}^{n_3} \cdots \sum_{i_d=1}^{n_d} \|(\boldsymbol{\mathcal{L}}_{A,\mathfrak{L}}(:,:,i_3,\ldots,i_d)\boldsymbol{\mathcal{R}}_{A,\mathfrak{L}}(:,:,i_3,\ldots,i_d)^T) \circ (\boldsymbol{\mathcal{L}}_{B,\mathfrak{L}}(:,:,i_3,\ldots,i_d)\boldsymbol{\mathcal{R}}_{B,\mathfrak{L}}(:,:,i_3,\ldots,i_d)^T)\|_*, \end{aligned}$$

where we denote $\boldsymbol{\mathcal{L}}_{A,\mathfrak{L}} = \mathfrak{L}(\boldsymbol{\mathcal{L}}_A)$ (similarly for $\boldsymbol{\mathcal{L}}_{B,\mathfrak{L}}$, $\boldsymbol{\mathcal{R}}_{A,\mathfrak{L}}$ and $\boldsymbol{\mathcal{R}}_{B,\mathfrak{L}}$), and we can decompose each term above into sum of rank one matrices as follows:

$$\begin{aligned} &(\boldsymbol{\mathcal{L}}_{A,\mathfrak{L}}(:,:,i_3,\ldots,i_d)\boldsymbol{\mathcal{R}}_{A,\mathfrak{L}}(:,:,i_3,\ldots,i_d)^T) \circ (\boldsymbol{\mathcal{L}}_{B,\mathfrak{L}}(:,:,i_3,\ldots,i_d)\boldsymbol{\mathcal{R}}_{B,\mathfrak{L}}(:,:,i_3,\ldots,i_d)^T) \\ =&\Big(\sum_{i_2=1}^{r} \boldsymbol{\mathcal{L}}_{A,\mathfrak{L}}(:,i_2,i_3,\ldots,i_d)\boldsymbol{\mathcal{R}}_{A,\mathfrak{L}}(:,i_2,i_3,\ldots,i_d)^T\Big) \circ \Big(\sum_{i_2=1}^{r} \boldsymbol{\mathcal{L}}_{B,\mathfrak{L}}(:,i_2,i_3,\ldots,i_d)\boldsymbol{\mathcal{R}}_{B,\mathfrak{L}}(:,i_2,i_3,\ldots,i_d)^T\Big) \\ =&\sum_{i_2=1}^{r}\sum_{i_2'=1}^{r} \big(\boldsymbol{\mathcal{L}}_{A,\mathfrak{L}}(:,i_2,i_3,\ldots,i_d) \circ \boldsymbol{\mathcal{L}}_{B,\mathfrak{L}}(:,i_2',i_3,\ldots,i_d)\big) \cdot \big(\boldsymbol{\mathcal{R}}_{A,\mathfrak{L}}(:,i_2,i_3,\ldots,i_d) \circ \boldsymbol{\mathcal{R}}_{B,\mathfrak{L}}(:,i_2',i_3,\ldots,i_d)\big)^T. \end{aligned}$$

So one can upper bound the nuclear norm via

$$
\begin{aligned}
&\|(\boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T) \circ (\boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T)\|_{\circledast,\mathfrak{L}} \\
\leq &\sum_{i_3=1}^{n_3} \cdots \sum_{i_d=1}^{n_d} \sum_{i_2=1}^{r} \sum_{i_2'=1}^{r} \frac{1}{\ell} \Big\| \big(\boldsymbol{\mathcal{L}}_{A,\mathfrak{L}}(:, i_2, i_3, \ldots, i_d) \circ \boldsymbol{\mathcal{L}}_{B,\mathfrak{L}}(:, i_2', i_3, \ldots, i_d)\big) \cdot \big(\boldsymbol{\mathcal{R}}_{A,\mathfrak{L}}(:, i_2, i_3, \ldots, i_d) \circ \boldsymbol{\mathcal{R}}_{B,\mathfrak{L}}(:, i_2', i_3, \ldots, i_d)\big)^T \Big\|_* \\
= &\sum_{i_3=1}^{n_3} \cdots \sum_{i_d=1}^{n_d} \sum_{i_2=1}^{r} \sum_{i_2'=1}^{r} \frac{1}{\ell} \Big\| \boldsymbol{\mathcal{L}}_{A,\mathfrak{L}}(:, i_2, i_3, \ldots, i_d) \circ \boldsymbol{\mathcal{L}}_{B,\mathfrak{L}}(:, i_2', i_3, \ldots, i_d) \Big\|_2 \Big\| \boldsymbol{\mathcal{R}}_{A,\mathfrak{L}}(:, i_2, i_3, \ldots, i_d) \circ \boldsymbol{\mathcal{R}}_{B,\mathfrak{L}}(:, i_2', i_3, \ldots, i_d) \Big\|_2 \\
= &\sum_{i_3=1}^{n_3} \cdots \sum_{i_d=1}^{n_d} \sum_{i_2=1}^{r} \sum_{i_2'=1}^{r} \frac{1}{\ell} \sqrt{\sum_{i_1=1}^{n_1} |(\boldsymbol{\mathcal{L}}_{A,\mathfrak{L}})_{i_1,i_2,i_3,\ldots,i_d}|^2 |(\boldsymbol{\mathcal{L}}_{B,\mathfrak{L}})_{i_1,i_2',i_3,\ldots,i_d}|^2} \sqrt{\sum_{i_1=1}^{n_2} |(\boldsymbol{\mathcal{R}}_{A,\mathfrak{L}})_{i_1,i_2,i_3,\ldots,i_d}|^2 |(\boldsymbol{\mathcal{R}}_{B,\mathfrak{L}})_{i_1,i_2',i_3,\ldots,i_d}|^2},
\end{aligned}
$$

where we replace nuclear norm by vector ℓ_2 norms in the third line because the summands are all rank one matrices. Now apply Cauchy-Schwarz inequality twice to obtain

$$
\begin{aligned}
&\|(\boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T) \circ (\boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T)\|_{\circledast,\mathfrak{L}} \\
\leq &\sum_{i_3=1}^{n_3} \cdots \sum_{i_d=1}^{n_d} \frac{1}{\ell} \sqrt{\sum_{i_2=1}^{r} \sum_{i_2'=1}^{r} \sum_{i_1=1}^{n_1} |(\boldsymbol{\mathcal{L}}_{A,\mathfrak{L}})_{i_1,i_2,i_3,\ldots,i_d}|^2 |(\boldsymbol{\mathcal{L}}_{B,\mathfrak{L}})_{i_1,i_2',i_3,\ldots,i_d}|^2} \\
&\qquad \sqrt{\sum_{i_2=1}^{r} \sum_{i_2'=1}^{r} \sum_{i_1=1}^{n_2} |(\boldsymbol{\mathcal{R}}_{A,\mathfrak{L}})_{i_1,i_2,i_3,\ldots,i_d}|^2 |(\boldsymbol{\mathcal{R}}_{B,\mathfrak{L}})_{i_1,i_2',i_3,\ldots,i_d}|^2} \\
= &\sum_{i_3=1}^{n_3} \cdots \sum_{i_d=1}^{n_d} \frac{1}{\ell} \sqrt{\sum_{i_1=1}^{n_1} \|\boldsymbol{\mathcal{L}}_{A,\mathfrak{L}}(i_1, :, i_3, \ldots, i_d)\|_2^2 \|\boldsymbol{\mathcal{L}}_{B,\mathfrak{L}}(i_1, :, i_3, \ldots, i_d)\|_2^2} \\
&\qquad \sqrt{\sum_{i_1=1}^{n_2} \|\boldsymbol{\mathcal{R}}_{A,\mathfrak{L}}(i_1, :, i_3, \ldots, i_d)\|_2^2 \|\boldsymbol{\mathcal{R}}_{B,\mathfrak{L}}(i_1, :, i_3, \ldots, i_d)\|_2^2} \\
\leq &\ell \sqrt{\sum_{i_3=1}^{n_3} \cdots \sum_{i_d=1}^{n_d} \sum_{i_1=1}^{n_1} \|\boldsymbol{\mathcal{L}}_A(i_1, :, i_3, \ldots, i_d)\|_2^2 \|\boldsymbol{\mathcal{L}}_B(i_1, :, i_3, \ldots, i_d)\|_2^2} \\
&\qquad \sqrt{\sum_{i_3=1}^{n_3} \cdots \sum_{i_d=1}^{n_d} \sum_{i_1=1}^{n_2} \|\boldsymbol{\mathcal{R}}_A(i_1, :, i_3, \ldots, i_d)\|_2^2 \|\boldsymbol{\mathcal{R}}_B(i_1, :, i_3, \ldots, i_d)\|_2^2} \\
\leq &\ell \Big(\|\boldsymbol{\mathcal{L}}_A\|_{2\diamond\infty} \|\boldsymbol{\mathcal{L}}_B\|_F \wedge \|\boldsymbol{\mathcal{L}}_A\|_F \|\boldsymbol{\mathcal{L}}_B\|_{2\diamond\infty} \Big) \Big(\|\boldsymbol{\mathcal{R}}_A\|_{2\diamond\infty} \|\boldsymbol{\mathcal{R}}_B\|_F \wedge \|\boldsymbol{\mathcal{R}}_A\|_F \|\boldsymbol{\mathcal{R}}_B\|_{2\diamond\infty} \Big).
\end{aligned} \tag{18}
$$

Putting (17) and (18) together, we have

$$
\begin{aligned}
&|p^{-1} \langle \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T), \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T) \rangle - \langle \boldsymbol{\mathcal{L}}_A *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_A^T, \boldsymbol{\mathcal{L}}_B *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_B^T \rangle| \\
\leq &c \ell^{\frac{3}{2}} \sqrt{\frac{n_{(1)} \log(n_{(1)} \ell)}{p}} \\
&\qquad \Big(\|\boldsymbol{\mathcal{L}}_A\|_{2\diamond\infty} \|\boldsymbol{\mathcal{L}}_B\|_F \wedge \|\boldsymbol{\mathcal{L}}_A\|_F \|\boldsymbol{\mathcal{L}}_B\|_{2\diamond\infty} \Big) \Big(\|\boldsymbol{\mathcal{R}}_A\|_{2\diamond\infty} \|\boldsymbol{\mathcal{R}}_B\|_F \wedge \|\boldsymbol{\mathcal{R}}_A\|_F \|\boldsymbol{\mathcal{R}}_B\|_{2\diamond\infty} \Big).
\end{aligned}
$$

□

Now we prove Lemma 4. First, define the event G as the intersection of the events that the bounds in Corollary 15 and Lemma 16 hold. The rest of the proof is under the assumption that G holds, which happens with high probability. By the condition $\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \leq \frac{0.02}{\sqrt{\ell}} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$ and Lemma 8, one knows that

$\boldsymbol{\mathcal{Q}}_t$ exists. For notational convenience, we denote $\boldsymbol{\mathcal{L}}_\sharp \coloneqq \boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{Q}}_t$, $\boldsymbol{\mathcal{R}}_\sharp \coloneqq \boldsymbol{\mathcal{R}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{Q}}_t^{-T}$, $\boldsymbol{\mathcal{L}}_\triangle \coloneqq \boldsymbol{\mathcal{L}}_\sharp - \boldsymbol{\mathcal{L}}_\star$, and $\boldsymbol{\mathcal{R}}_\triangle \coloneqq \boldsymbol{\mathcal{R}}_\sharp - \boldsymbol{\mathcal{R}}_\star$, and $\epsilon \coloneqq 0.02$. In addition, denote $\widetilde{\boldsymbol{\mathcal{F}}}_{t+1}$ as the update before projection as

$$\widetilde{\boldsymbol{\mathcal{F}}}_{t+1} \coloneqq \begin{bmatrix} \widetilde{\boldsymbol{\mathcal{L}}}_{t+1} \\ \widetilde{\boldsymbol{\mathcal{R}}}_{t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mathcal{L}}_t - \frac{\eta}{p}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t^T - \boldsymbol{\mathcal{X}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_t^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t)^{-1} \\ \boldsymbol{\mathcal{R}}_t - \frac{\eta}{p}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t^T - \boldsymbol{\mathcal{X}}_\star)^T *_{\mathfrak{L}} \boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} (\boldsymbol{\mathcal{L}}_t^T *_{\mathfrak{L}} \boldsymbol{\mathcal{L}}_t)^{-1} \end{bmatrix},$$

and therefore $\boldsymbol{\mathcal{F}}_{t+1} = \mathcal{P}_B(\widetilde{\boldsymbol{\mathcal{F}}}_{t+1})$. Note that in view of Lemma 3, it suffices to prove the following relation

$$\mathrm{dist}(\widetilde{\boldsymbol{\mathcal{F}}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) \le (1 - 0.6\eta)\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \tag{19}$$

because the second conclusion is a simple consequence of Lemma 10 as

$$\begin{aligned} \|\boldsymbol{\mathcal{L}}_t *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_t^T - \boldsymbol{\mathcal{X}}_\star\|_F &\le (1+\epsilon)(\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F + \|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F) \\ &\le (1+\epsilon)\sqrt{2}\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \\ &\le 1.5\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star), \end{aligned} \tag{20}$$

where we used $a + b \le \sqrt{2(a^2+b^2)}$ in the second row. In what follows, we focus on proving (19). We begin by listing a few easy consequences under the assumed conditions.

Lemma 17. *Under conditions* $\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \le \frac{\epsilon}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$ *and* $\sqrt{n_1}\|\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T\|_{2,\infty} \vee \sqrt{n_2}\|\boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{L}}_\sharp^T\|_{2,\infty} \le C_B\sqrt{\frac{\mu r}{\ell}}\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star)$*, we have*

$$\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\| \vee \|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\| \le \epsilon; \tag{21a}$$

$$\|\boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\| \le \frac{1}{1-\epsilon}; \tag{21b}$$

$$\|\boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\| \le \frac{1}{(1-\epsilon)^2}; \tag{21c}$$

$$\sqrt{n_1}\|\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|\boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_{2,\infty} \le \frac{C_B}{1-\epsilon}\sqrt{\frac{\mu r}{\ell}}\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star); \tag{21d}$$

$$\sqrt{n_1}\|\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|\boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|_{2,\infty} \le \frac{C_B\kappa}{1-\epsilon}\sqrt{\frac{\mu r}{\ell}}; \tag{21e}$$

$$\sqrt{n_1}\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_{2,\infty} \le (1 + \frac{C_B}{1-\epsilon})\sqrt{\frac{\mu r}{\ell}}\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star). \tag{21f}$$

Now we are ready to prove (19), which follows the exact same steps as in Lemma 5 of Wu (2025). By the definition of $\mathrm{dist}(\widetilde{\boldsymbol{\mathcal{F}}}_{t+1}, \boldsymbol{\mathcal{F}}_\star)$, we have

$$\mathrm{dist}^2(\widetilde{\boldsymbol{\mathcal{F}}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) \le \|(\widetilde{\boldsymbol{\mathcal{L}}}_{t+1} *_{\mathfrak{L}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 + \|(\widetilde{\boldsymbol{\mathcal{R}}}_{t+1} *_{\mathfrak{L}} \boldsymbol{\mathcal{Q}}_t^{-T} - \boldsymbol{\mathcal{R}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2. \tag{22}$$

Plugging in the update rule (9) and the decomposition $\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T - \boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T + \boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T$ to obtain

$$\begin{aligned} &(\widetilde{\boldsymbol{\mathcal{L}}}_{t+1} *_{\mathfrak{L}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} \\ =&\Big(\boldsymbol{\mathcal{L}}_\sharp - \eta p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T - \boldsymbol{\mathcal{X}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} - \boldsymbol{\mathcal{L}}_\star\Big) *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} \\ =&(1-\eta)\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} - \eta\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} \\ &\quad - \eta(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T - \boldsymbol{\mathcal{X}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}. \end{aligned}$$

This allows us to expand the square of the first term in (22) as

$$
\begin{aligned}
&\|(\widetilde{\boldsymbol{\mathcal{L}}}_{t+1} *_{\mathfrak{L}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 \\
=&\|(1-\eta)\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} - \eta \boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 \\
&- 2\eta(1-\eta)\langle \boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}, (p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T - \boldsymbol{\mathcal{X}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\rangle \\
&+ 2\eta^2 \langle \boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}, \\
&\qquad\qquad (p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T - \boldsymbol{\mathcal{X}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\rangle \\
&+ \eta^2 \|(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T - \boldsymbol{\mathcal{X}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 \\
:=&\mathfrak{P}_1 - \mathfrak{P}_2 + \mathfrak{P}_3 + \mathfrak{P}_4.
\end{aligned}
$$

Bound of $\mathfrak{P}_1$. This term can be controlled as Equation 38 in Wu (2025) as follows.

$$
\mathfrak{P}_1 \le \Big((1-\eta)^2 + \frac{2\epsilon}{1-\epsilon}\eta(1-\eta)\Big)\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 + \frac{2\epsilon+\epsilon^2}{(1-\epsilon)^2}\eta^2\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2.
$$

Bound of $\mathfrak{P}_2$. Using the decomposition $\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T - \boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T + \boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T$ and applying the triangle inequality to obtain

$$
\begin{aligned}
|\mathfrak{P}_2| &= 2\eta(1-\eta)\Big|\langle \boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}, (p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T - \boldsymbol{\mathcal{X}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\rangle\Big| \\
&\le 2\eta(1-\eta)\Big(\Big|\langle \boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}, (p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\rangle\Big| \\
&\quad + \Big|\langle \boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}, (p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\rangle\Big| \\
&\quad + \Big|\langle \boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}, (p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\rangle\Big|\Big) \\
&:= 2\eta(1-\eta)(\mathfrak{P}_{2,1} + \mathfrak{P}_{2,2} + \mathfrak{P}_{2,3}).
\end{aligned}
$$

For the first term $\mathfrak{P}_{2,1}$, we can invoke Corollary 15 to obtain

$$
\begin{aligned}
\mathfrak{P}_{2,1} &\le C_1\sqrt{\frac{\mu r \log(n_{(1)}\ell)}{p n_{(2)}\ell}}\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T\|_F \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T\|_F \\
&\le C_1\sqrt{\frac{\mu r \log(n_{(1)}\ell)}{p n_{(2)}\ell}}\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T\| \\
&\qquad\qquad \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\| \|\boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T\| \\
&= C_1\sqrt{\frac{\mu r \log(n_{(1)}\ell)}{p n_{(2)}\ell}}\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 \|\boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\| \\
&\le \frac{C_1}{(1-\epsilon)^2}\sqrt{\frac{\mu r \log(n_{(1)}\ell)}{p n_{(2)}\ell}}\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2,
\end{aligned}
$$

where the last inequality uses (21c). For the first term $\mathfrak{P}_{2,2}$, we can invoke Lemma 16 with $\boldsymbol{\mathcal{L}}_A \coloneqq \boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}$, $\boldsymbol{\mathcal{R}}_A \coloneqq \boldsymbol{\mathcal{R}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}$, $\boldsymbol{\mathcal{L}}_B \coloneqq \boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}$, $\boldsymbol{\mathcal{R}}_B \coloneqq \boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}$ to obtain

$$
\begin{aligned}
\mathfrak{P}_{2,2} &\le C_2 \ell^{\frac{3}{2}} \sqrt{\frac{n_{(1)} \log(n_{(1)}\ell)}{p}} \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_{2\diamond\infty} \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \\
&\qquad \|\boldsymbol{\mathcal{R}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|_{2\diamond\infty} \|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \\
&\le C_2 \ell^{\frac{3}{2}} \sqrt{\frac{n_{(1)} \log(n_{(1)}\ell)}{p}} \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \\
&\qquad \|\boldsymbol{\mathcal{R}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|.
\end{aligned}
$$

Similarly, we can bound $\mathfrak{P}_{2,3}$ as

$$
\begin{aligned}
\mathfrak{P}_{2,3} &\le C_2 \ell^{\frac{3}{2}} \sqrt{\frac{n_{(1)} \log(n_{(1)}\ell)}{p}} \|\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \\
&\qquad \|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|.
\end{aligned}
$$

Utilizing the consequences in Lemma 17, we have

$$
\begin{aligned}
\mathfrak{P}_{2,2} &\le \frac{C_2 \mu r \kappa}{(1-\epsilon)^2} \Big(1 + \frac{C_B}{1-\epsilon}\Big) \sqrt{\frac{\ell \log(n_{(1)}\ell)}{n_{(2)} p}} \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F; \\
\mathfrak{P}_{2,3} &\le \frac{C_2 C_B^2 \mu r \kappa^2}{(1-\epsilon)^4} \sqrt{\frac{\ell \log(n_{(1)}\ell)}{n_{(2)} p}} \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F.
\end{aligned}
$$

We then combine the bounds for $\mathfrak{P}_{2,1}$, $\mathfrak{P}_{2,2}$ and $\mathfrak{P}_{2,3}$ to arrive at

$$
\begin{aligned}
\mathfrak{P}_2 &\le 2\eta(1-\eta) \Big(\frac{C_1}{(1-\epsilon)^2} \sqrt{\frac{\mu r \log(n_{(1)}\ell)}{p n_{(2)} \ell}} \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 \\
&\qquad + \frac{C_2 \mu r \kappa}{(1-\epsilon)^2} \Big(1 + \frac{C_B}{1-\epsilon} + \frac{C_B^2 \kappa}{(1-\epsilon)^2}\Big) \sqrt{\frac{\ell \log(n_{(1)}\ell)}{n_{(2)} p}} \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \Big) \\
&= 2\eta(1-\eta) \Big(\nu_1 \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 + \nu_2 \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \Big) \\
&\le \eta(1-\eta) \Big((2\nu_1 + \nu_2) \|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 + \nu_2 \|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 \Big),
\end{aligned}
$$

where we denote

$$
\nu_1 \coloneqq \frac{C_1}{(1-\epsilon)^2} \sqrt{\frac{\mu r \log(n_{(1)}\ell)}{p n_{(2)} \ell}} \quad \text{and} \quad \nu_2 \coloneqq \frac{C_2 \mu r \kappa}{(1-\epsilon)^2} \Big(1 + \frac{C_B}{1-\epsilon} + \frac{C_B^2 \kappa}{(1-\epsilon)^2}\Big) \sqrt{\frac{\ell \log(n_{(1)}\ell)}{n_{(2)} p}}.
$$

Bound of $\mathfrak{P}_3$. For the term $\mathfrak{P}_3$, we first have

$$
\begin{aligned}
|\mathfrak{P}_3| &\le 2\eta^2 \Big(\Big| \langle \boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}, \\
&\qquad (p^{-1} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} \rangle \Big| \\
&\quad + \Big| \langle \boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}, \\
&\qquad (p^{-1} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} \rangle \Big| \Big) \\
&\coloneqq 2\eta^2 (\mathfrak{P}_{3,1} + \mathfrak{P}_{3,2}).
\end{aligned}
$$

We invoke Lemma 16 to bound $\mathfrak{P}_{3,1}$ as

$$\begin{aligned}
\mathfrak{P}_{3,1} &\le C_2\ell^{\frac{3}{2}}\sqrt{\frac{n_{(1)}\log(n_{(1)}\ell)}{p}}\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F\|\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|_{2\diamond\infty} \\
&\qquad \|\boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|_{2\diamond\infty}\|\boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \\
&\le C_2\ell^{\frac{3}{2}}\sqrt{\frac{n_{(1)}\log(n_{(1)}\ell)}{p}}\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F\|\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|_{2,\infty} \\
&\qquad \|\boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{-\frac{1}{2}}\|_{2,\infty}\|\boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|^2\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \\
&\le \frac{C_2C_B\mu r\kappa}{(1-\epsilon)^3}\sqrt{\frac{\ell\log(n_{(1)}\ell)}{n_{(2)}p}}\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F.
\end{aligned}$$

For $\mathfrak{P}_{3,2}$, we again invoke Corollary 15 to obtain

$$\begin{aligned}
\mathfrak{P}_{3,2} &\le C_1\sqrt{\frac{\mu r\log(n_{(1)}\ell)}{pn_{(2)}\ell}}\|\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T\|_F\|\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T\|_F \\
&\le C_1\sqrt{\frac{\mu r\log(n_{(1)}\ell)}{pn_{(2)}\ell}}\|\boldsymbol{\mathcal{L}}_\star *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T\|_F^2\|\boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|^2 \\
&\le \frac{C_1}{(1-\epsilon)^2}\sqrt{\frac{\mu r\log(n_{(1)}\ell)}{pn_{(2)}\ell}}\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2.
\end{aligned}$$

Combining the bounds for $\mathfrak{P}_{3,1}$ and $\mathfrak{P}_{3,2}$ to arrive at

$$\begin{aligned}
|\mathfrak{P}_3| &\le 2\eta^2\Big(\frac{C_1}{(1-\epsilon)^2}\sqrt{\frac{\mu r\log(n_{(1)}\ell)}{pn_{(2)}\ell}}\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 + \frac{C_2C_B\mu r\kappa}{(1-\epsilon)^3}\sqrt{\frac{\ell\log(n_{(1)}\ell)}{n_{(2)}p}}\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F\Big) \\
&\le 2\eta^2\Big(\nu_1\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 + \nu_2\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F\Big) \\
&\le \eta^2\Big(\nu_2\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F^2 + (2\nu_1+\nu_2)\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F\Big).
\end{aligned}$$

Bound of $\mathfrak{P}_4$. Moving to the term $\mathfrak{P}_4$, we have

$$\begin{aligned}
\sqrt{\mathfrak{P}_4} &= \eta\|(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T - \boldsymbol{\mathcal{X}}_\star) *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F \\
&\le \eta\Big(\Big|\langle(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T), \widetilde{\boldsymbol{\mathcal{L}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T\rangle\Big| \\
&\qquad + \Big|\langle(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\star^T), \widetilde{\boldsymbol{\mathcal{L}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T\rangle\Big| \\
&\qquad + \Big|\langle(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\triangle^T), \widetilde{\boldsymbol{\mathcal{L}}} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}} *_{\mathfrak{L}} (\boldsymbol{\mathcal{R}}_\sharp^T *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\mathfrak{L}} \boldsymbol{\mathcal{R}}_\sharp^T\rangle\Big| \\
&:= \eta(\mathfrak{P}_{4,1} + \mathfrak{P}_{4,2} + \mathfrak{P}_{4,3}),
\end{aligned}$$

where we have used the variational representation of the Frobenius norm for some $\widetilde{\boldsymbol{\mathcal{L}}} \in \mathbb{R}^{n_1\times r\times n_3\times\cdots\times n_d}$ obeying $\|\widetilde{\boldsymbol{\mathcal{L}}}\|_F \le 1$. Note that the decomposition of $\sqrt{\mathfrak{P}_4}$ is extremely similar to that of $\mathfrak{P}_2$, thus we can follow a similar argument to control these terms as

$$\begin{aligned}
\mathfrak{P}_{4,1} &\le \frac{C_1}{(1-\epsilon)^2}\sqrt{\frac{\mu r\log(n_{(1)}\ell)}{pn_{(2)}\ell}}\|\boldsymbol{\mathcal{L}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F; \\
\mathfrak{P}_{4,2} &\le \frac{C_2\mu r\kappa}{(1-\epsilon)^2}\Big(1+\frac{C_B}{1-\epsilon}\Big)\sqrt{\frac{\ell\log(n_{(1)}\ell)}{n_{(2)}p}}\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F; \\
\mathfrak{P}_{4,3} &\le \frac{C_2C_B^2\mu r\kappa^2}{(1-\epsilon)^4}\sqrt{\frac{\ell\log(n_{(1)}\ell)}{n_{(2)}p}}\|\boldsymbol{\mathcal{R}}_\triangle *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_\star^{\frac{1}{2}}\|_F.
\end{aligned}$$

Hence,

$$\sqrt{\mathfrak{P}_4} \leq \eta(\nu_1 \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_{\star}^{\frac{1}{2}}\|_F + \nu_2 \|\boldsymbol{\mathcal{R}}_{\triangle} *_{\mathfrak{L}} \boldsymbol{\mathcal{S}}_{\star}^{\frac{1}{2}}\|_F).$$

We omit the rest of the proof since it is identical to the one for Lemma 5 of Wu (2025).

A.4 Proof of Lemma 5

Following the proof of Lemma 6 in Wu (2025), we have

$$\operatorname{dist}(\widetilde{\boldsymbol{\mathcal{F}}}_0, \boldsymbol{\mathcal{F}}_\star) \leq 5\sqrt{\frac{s_r}{\ell}} \|p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star) - \boldsymbol{\mathcal{X}}_\star\|.$$

Using Lemma 12, we know that

$$\|(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{X}}_\star)\| \leq c\Big(\frac{\log(n_{(1)}\ell)}{p}\|\boldsymbol{\mathcal{X}}_\star\|_\infty + \sqrt{\frac{\log(n_{(1)}\ell)}{p}}\|\boldsymbol{\mathcal{X}}_\star\|_{\infty,2}\Big),$$

holds with high probability. The proof is finished by applying Lemma 6 and Lemma 7 and plugging the following bounds from incoherence assumption of $\boldsymbol{X}_\star$:

$$\|\boldsymbol{\mathcal{X}}_\star\|_\infty \leq \sqrt{\ell}\|\boldsymbol{\mathcal{U}}_\star\|_{2,\infty}\|\boldsymbol{\mathcal{S}}_\star\|\|\boldsymbol{\mathcal{V}}_\star\|_{2,\infty} \leq \frac{\mu r}{\sqrt{n_1 n_2 \ell}}\kappa\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star);$$

$$\|\boldsymbol{\mathcal{X}}_\star\|_{\infty,2} \leq \|\boldsymbol{\mathcal{U}}_\star\|_{2,\infty}\|\boldsymbol{\mathcal{S}}_\star\|\|\boldsymbol{\mathcal{V}}_\star\| \vee \|\boldsymbol{\mathcal{U}}_\star\|\|\boldsymbol{\mathcal{S}}_\star\|\|\boldsymbol{\mathcal{V}}_\star\|_{2,\infty} \leq \sqrt{\frac{\mu r}{n_{(2)}\ell}}\kappa\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).$$