Guaranteed Nonconvex Low-Rank Tensor Estimation via Scaled Gradient Descent

Tong Wu   WUTONG@BIGAI.AI

Beijing Institute for General Artificial Intelligence
Beijing 100080, China

Editor: Qiang Liu

Abstract

Tensors, which give a faithful and effective representation to deliver the intrinsic structure of multi-dimensional data, play a crucial role in an increasing number of signal processing and machine learning problems. However, tensor data are often accompanied by arbitrary signal corruptions, including missing entries and sparse noise. A fundamental challenge is to reliably extract the meaningful information from corrupted tensor data in a statistically and computationally efficient manner. This paper develops a scaled gradient descent (ScaledGD) algorithm to directly estimate the tensor factors with tailored spectral initializations under the tensor-tensor product (t-product) and tensor singular value decomposition (t-SVD) framework. With tailored variants for tensor robust principal component analysis, (robust) tensor completion and tensor regression, we theoretically show that ScaledGD achieves linear convergence at a constant rate that is independent of the condition number of the ground truth low-rank tensor, while maintaining the low per-iteration cost of gradient descent. To the best of our knowledge, ScaledGD is the first algorithm that provably has such properties for low-rank tensor estimation with the t-SVD. Finally, numerical examples are provided to demonstrate the efficacy of ScaledGD in accelerating the convergence rate of ill-conditioned low-rank tensor estimation in a number of applications. Our code is available at `https://github.com/twugithub/Tensor-ScaledGD`.

Keywords: Low-rank tensor factorization, robust tensor completion, scaled gradient descent, tensor completion, tensor regression, tensor robust PCA, tensor singular value decomposition

1. Introduction

With the increasing availability of high-dimensional and multiway datasets, we are witnessing a growing demand for efficient data analysis techniques. Many practical applications collect data that are naturally in the form of a tensor. Instances of tensor data include images, videos, hyperspectral images, and signals generated by magnetic resonance systems. Tensors arise naturally as a powerful model for better capturing the multiway relationships and interactions within data than matrices; examples include image processing (Liu et al., 2013), climate forecasting (Yu et al., 2015), topic modeling (Anandkumar et al., 2014), and neuroimaging data analysis (Ahmed et al., 2020). In this work, specifically, we consider the tensor estimation problem, which is the central task across many problems. Mathematically, the goal of tensor estimation is to estimate a K-way tensor $\boldsymbol{\mathcal{X}}_\star \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ from a

set of observations $\boldsymbol{y} \in \mathbb{R}^m$ given by

$$\boldsymbol{y} \approx \mathcal{A}(\boldsymbol{\mathcal{X}}_\star), \tag{1}$$

where $\mathcal{A}(\cdot)$ is a linear map that models the data collection process. For ease of presentation, we consider the case $K = 3$ throughout the paper, while the general case will be investigated in future work. Arguably, the data of interest can be well represented by a much smaller number of latent factors compared to the dimensionality of the ambient space, which suggests exploiting low-dimensional geometric structures for meaningful recovery.

1.1 Low-rank Tensor Estimation

One of the most widely adopted low-rank tensor decompositions—which is the focus of this paper—considers low-rank structure under the *tensor singular value decomposition (t-SVD)* (Kilmer et al., 2013). Specifically, we assume the ground truth tensor $\boldsymbol{\mathcal{X}}_\star$ of tensor tubal rank-r that admits the following t-SVD:

$$\boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{U}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}_\star^H,$$

where $*_{\boldsymbol{\Phi}}$ denotes the t-product evoked by the transformation matrix $\boldsymbol{\Phi}$, $\boldsymbol{\mathcal{U}}_\star \in \mathbb{R}^{n_1 \times r \times n_3}$ and $\boldsymbol{\mathcal{V}}_\star \in \mathbb{R}^{n_2 \times r \times n_3}$ are orthogonal tensors, and $\boldsymbol{\mathcal{G}}_\star \in \mathbb{R}^{r \times r \times n_3}$ is an f-diagonal tensor (see Definitions in Section 2.1). Compared with other tensor decomposition strategies such as CANDECOMP/PARAFAC (CP) decomposition (Kiers, 2000) and Tucker decomposition (Tucker, 1966), t-SVD can take advantage of structures in both the original and transformed domains. It has been observed that after conducting Discrete Fourier Transform (DFT) along the third dimension of a three-way tensor, the transformed tensor may exhibit low-rank structure in the Fourier domain (Zhang et al., 2014), which can be characterized by the tensor tubal rank and tensor nuclear norm (Kilmer et al., 2013). Low-rankness in the spectral domain extends traditional models that require strong low-rankness in the original domain. The t-SVD based models evoked by the tubal rank own the same tight recovery bound as the matrix cases (Lu et al., 2018). Another advantage of the t-SVD scheme is its superior capability in capturing the "spatial-shifting" correlation that is ubiquitous in real multiway data. Interestingly, the t-SVD can be generalized by replacing DFT with any invertible linear transforms (Kernfeld et al., 2015). Driven by these advantages, extensive numerical examples have demonstrated its efficacy in many applications (Zhang and Aeron, 2017; Lu et al., 2018, 2019; Song et al., 2020; Kilmer et al., 2021; Zhou et al., 2021; Kong et al., 2021; Lu, 2021; Qin et al., 2022; Wu, 2026).

Motivating examples. We focus on optimization problems that look for low-rank tensors using partial or corrupted observations.

- *Tensor RPCA.* We first consider the tensor robust principal component analysis (RPCA) problem, which attempts to find a low-rank tensor that best approximates grossly corrupted observations. Mathematically, imagine that we are given a data tensor

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}_\star + \boldsymbol{\mathcal{S}}_\star, \tag{2}$$

where $\boldsymbol{\mathcal{X}}_\star$ is low-rank and $\boldsymbol{\mathcal{S}}_\star$ is sparse, and both components are of arbitrary magnitudes. Our goal is to recover $\boldsymbol{\mathcal{X}}_\star$ and $\boldsymbol{\mathcal{S}}_\star$ from the corrupted observation $\boldsymbol{\mathcal{Y}}$ in an efficient manner.

- *Tensor completion.* We then study the low-rank tensor completion problem, which aims to recover a low-rank tensor from only a small subset of its revealed observations. Mathematically, we are given entries

$$[\boldsymbol{\mathcal{X}}_\star]_{i,j,k}, \quad (i,j,k) \in \boldsymbol{\Omega},$$

where $\boldsymbol{\Omega}$ denotes the set of the indices of the observed entries. The goal is to recover $\boldsymbol{\mathcal{X}}_\star$ from these observed entries.

- *Robust tensor completion (RTC).* In this problem, we attempt to recover an underlying low-rank tensor by observing a small number of sparsely corrupted entries. Formally, let $\boldsymbol{\mathcal{Y}}$ be a tensor with a decomposition $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}_\star + \boldsymbol{\mathcal{S}}_\star$, where $\boldsymbol{\mathcal{X}}_\star$ is low-rank and $\boldsymbol{\mathcal{S}}_\star$ is a sparse corruption tensor whose few non-zero entries are arbitrary. The RTC problem is to recover $\boldsymbol{\mathcal{X}}_\star$ from the observed entries $\{\boldsymbol{\mathcal{Y}}_{i,j,k} | (i,j,k) \in \boldsymbol{\Omega}\}$, where $\boldsymbol{\Omega}$ denotes the locations of the observed entries in $\boldsymbol{\mathcal{Y}}$.

- *Tensor regression.* We finally consider a linear tensor-structured regression model, where the i-th response or observation is given by

$$y_i = \langle \boldsymbol{\mathcal{A}}_i, \boldsymbol{\mathcal{X}}_\star \rangle, \quad i = 1, 2, \ldots, m,$$

where $\boldsymbol{\mathcal{A}}_i \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the i-th measurement tensor and $\langle \cdot, \cdot \rangle$ denotes the canonical inner product. The goal is to recover the low-rank tensor $\boldsymbol{\mathcal{X}}_\star$ from the responses $\boldsymbol{y} = \{y_i\}_{i=1}^m$.

1.2 Main Contributions

In view of the low-rank t-SVD, a natural approach to estimate the low-rank tensor $\boldsymbol{\mathcal{X}}_\star$ in (1) is to parameterize $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}^H$ by two factor tensors $\boldsymbol{\mathcal{L}} \in \mathbb{R}^{n_1 \times r \times n_3}$ and $\boldsymbol{\mathcal{R}} \in \mathbb{R}^{n_2 \times r \times n_3}$, and then to estimate the ground truth factors via optimizing the unconstrained least-squares loss function:

$$\min_{\boldsymbol{\mathcal{L}} \in \mathbb{R}^{n_1 \times r \times n_3}, \boldsymbol{\mathcal{R}} \in \mathbb{R}^{n_2 \times r \times n_3}} f(\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{R}}) \coloneqq \|\mathcal{A}(\boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}^H) - \boldsymbol{y}\|_2^2, \tag{3}$$

where $\boldsymbol{\mathcal{R}}^H$ denotes the conjugate transpose of $\boldsymbol{\mathcal{R}}$ (see Definition 2) and we use the vector ℓ_2 norm to quantify the error. Despite the nonconvexity of the objective function, a simple and intuitive approach is to update the tensor factors via gradient descent. While remarkable progress has been made in recent years in the matrix setting (Chi et al., 2019), this line of research still remains relatively unexplored for the tensor setting, especially when it comes to provable sample and computational guarantees.

Motivated by the recent success of scaled gradient descent (ScaledGD) (Tong et al., 2021b,a, 2022; Dong et al., 2023b) for accelerating ill-conditioned low-rank estimation, we propose to update the tensor factors iteratively along the scaled gradient directions:

$$\begin{aligned} \boldsymbol{\mathcal{L}}_{t+1} &= \boldsymbol{\mathcal{L}}_t - \eta \nabla_{\boldsymbol{\mathcal{L}}} f(\boldsymbol{\mathcal{L}}_t, \boldsymbol{\mathcal{R}}_t) *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_t^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t)^{-1}, \\ \boldsymbol{\mathcal{R}}_{t+1} &= \boldsymbol{\mathcal{R}}_t - \eta \nabla_{\boldsymbol{\mathcal{R}}} f(\boldsymbol{\mathcal{L}}_t, \boldsymbol{\mathcal{R}}_t) *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_t^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_t)^{-1}, \end{aligned} \tag{4}$$

where $\eta > 0$ is the learning rate and $\nabla_{\mathcal{L}} f(\mathcal{L}_t, \mathcal{R}_t)$ (resp., $\nabla_{\mathcal{R}} f(\mathcal{L}_t, \mathcal{R}_t)$) denotes the gradient of f with respect to $\mathcal{L}_t$ (resp., $\mathcal{R}_t$) at the t-th iteration. Compared with vanilla gradient descent, our method scales or preconditions the search directions of $\mathcal{L}_t$ and $\mathcal{R}_t$ in (4) by $(\mathcal{R}_t^H *_{\Phi} \mathcal{R}_t)^{-1}$ and $(\mathcal{L}_t^H *_{\Phi} \mathcal{L}_t)^{-1}$, respectively. With the preconditioners, ScaledGD updates the tensor factors with better search directions and larger step sizes. As the preconditioners are computed by inverting two $r \times r \times n_3$ tensors, whose size is much smaller than the dimension of the tensor factors, each iteration of ScaledGD only adds minimal overhead to the gradient computation, while maintaining a linear rate of convergence regardless of the condition number.

Theoretical guarantees. We investigate the theoretical properties of ScaledGD, which are notably more challenging than the matrix counterpart. Our model is more generic as it is allowed to use any invertible linear transforms that satisfy certain conditions. We establish that ScaledGD—when initialized with a tailored spectral initialization scheme—can achieve linear convergence at a rate *independent* of the condition number of the ground truth tensor with near-optimal sample complexities. To be concrete, we have the following guarantees:

- *Tensor RPCA.* Under the deterministic corruption model (Lu et al., 2020), ScaledGD converges linearly to the true low-rank tensor in both the Frobenius norm and the entrywise ℓ_∞ norm, as long as the level of corruptions—measured in terms of the fraction of nonzero entries per tube in each mode—does not exceed the order of $\min\left(\frac{n_3\sqrt{n_{(2)}n_2}}{\mu s_r^{1.5}\kappa(n_1+n_2n_3)}, \frac{n_3}{\mu^2 s_r^2 \kappa^2}\right)$, where $n_{(2)} = \min(n_1, n_2)$, μ and κ are the incoherence parameter and the condition number of the ground truth tensor $\mathcal{X}_\star$, respectively, and s_r is the summation of all the entries in the multi-rank of $\mathcal{X}_\star$ (to be formally defined later).

- *Tensor completion.* Under the Bernoulli sampling model, ScaledGD combined with a properly designed scaled projection step reaches ϵ-accuracy in $\mathcal{O}(\log(1/\epsilon))$ iterations, as long as the sample complexity satisfies $p \gtrsim \max\left(\frac{\mu s_r(n_1+n_2)\log(\max(n_1,n_2)n_3)}{n_1 n_2 n_3}, \frac{\mu^2 s_r^2 \kappa^4 \ell \log(\max(n_1,n_2)n_3)}{\min(n_1,n_2)n_3^2}\right)$, where ℓ is a constant related to the transformation matrix Φ.

- *Robust tensor completion.* Under the random Bernoulli observation model, ScaledGD succeeds with high probability as long as the fraction α of corruptions per tube for each mode satisfies $\alpha \lesssim \min\left(\frac{pn_3\sqrt{n_{(2)}n_2}}{\mu s_r^{1.5}\kappa(n_1+n_2n_3)}, \frac{pn_3}{\mu^2 s_r^2 \kappa^2}\right)$ and the probability of observation satisfies $p \gtrsim \log(\max(n_1,n_2)n_3)\max\left(\frac{\mu^{1.5}s_r^{1.5}(n_1+n_2)}{n_3\sqrt{n_1 n_2}}, \frac{\mu^2 s_r^2 \ell}{n_3\sqrt{\min(n_1,n_2)n_3}}, \frac{\mu s_r \kappa^2}{\sqrt{\min(n_1,n_2)n_3}}\right)$. Note that the recent work Cai et al. (2026) proposed a scalable method for robust matrix completion based on ScaledGD, the recovery guarantee derived in Cai et al. (2026) is under the assumption that the matrix is fully observed ($p = 1$). Remarkably, our analysis extends this result to the tensor case and considers the general case where the tensor is only partially observed.

- *Tensor regression.* As long as the measurement operator satisfies the tensor restricted isometry property (TRIP) (Rauhut et al., 2017) with a TRIP constant $\delta_{2r} \lesssim 1/(\sqrt{\frac{s_r}{\ell}}\kappa)$, ScaledGD reaches ϵ-accuracy in $\mathcal{O}(\log(1/\epsilon))$ iterations when initialized by the spectral method.

1.3 Related Work

Scaled first-order methods for low-rank matrix recovery. Variants of the scaled gradient methods have been proposed for low-rank matrix recovery (Mishra et al., 2012; Mishra and Sepulchre, 2016; Tanner and Wei, 2016; Tong et al., 2021a) that aim to approximate the low-rank matrix by the product of two factor matrices, where strong statistical guarantees were first established in Tong et al. (2021a). The matrix RPCA method in Cai et al. (2021) extended Tong et al. (2021a) by using a threshold-based trimming procedure to identify the sparse matrix. Jia et al. (2023) proved the global convergence of ScaledGD and alternating scaled gradient descent (AltScaledGD) for the nonconvex low-rank matrix factorization problem. There have been many other efforts in the literature to solve the low-rank matrix recovery problem with provable nonconvex optimization procedures; see, e.g., Chen and Wainwright (2015); Gu et al. (2016); Yi et al. (2016); Zheng and Lafferty (2016); Ge et al. (2017); Cherapanamjeri et al. (2017); Du et al. (2018); Zeng and So (2018); Li et al. (2019); Ye and Du (2021) for an incomplete list.

Low-rank tensor estimation using t-SVD. Turning to the tensor case, unfolding-based approaches typically lead to performance degradation since they neglect the high-order interactions within tensor data. Recently, motivated by the notion of tensor-tensor product (t-product) (Kilmer et al., 2013) and t-SVD scheme, a new tensor nuclear norm was proposed and applied in tensor RPCA (Lu et al., 2020; Gao et al., 2021; Lu, 2021), tensor completion (Zhang and Aeron, 2017; Lu et al., 2018, 2019; Song et al., 2020; Qin et al., 2022), and tensor data clustering (Zhou et al., 2021; Wu, 2024, 2026). To accelerate tensor RPCA, Qiu et al. (2022) proposed two alternating projection algorithms that exhibit linear convergence behavior. While this paper and Qiu et al. (2022) share the same tensor RPCA setup, our work is fundamentally different from Qiu et al. (2022). The methods proposed in Qiu et al. (2022) are based on the idea of alternating projection, which can be considered an extension of Netrapalli et al. (2014); Cai et al. (2019) in the matrix case to the case of tensors. In contrast, our work draws inspiration from Tong et al. (2021a); Cai et al. (2021) and parameterizes the low-rank term in (2) by two low-rank factors. To the best of our knowledge, this paper is the first work that provides rigorous statistical and computational guarantees for scaled gradient methods based on the t-SVD framework. There are many technical novelty in our analysis compared to Tong et al. (2021a); Cai et al. (2021). Specifically, the optimization problems in Tong et al. (2021a); Cai et al. (2021) primarily involve matrix-valued variables, whereas our optimization problem primarily involves tensor-valued variables. Because of this reason, all the inequalities involving matrices in Tong et al. (2021a); Cai et al. (2021) must be proved using some tensor properties involving t-product in our analysis.

Provable low-rank tensor estimation using other decompositions. As mentioned earlier, it is not straightforward to generalize low-rank matrix estimation methods to tensors because a tensor can be decomposed in many ways. Beyond the t-SVD, common tensor decompositions include CP (Kiers, 2000), Tucker (Tucker, 1966), HOSVD (Lathauwer et al., 2000), and tensor-train (Oseledets, 2011). Several works for low-rank tensor estimation relying on these decompositions have been proposed (Goldfarb and Qin, 2014; Anandkumar et al., 2016; Driggs et al., 2019; Xia and Yuan, 2019; Yuan et al., 2019; Yang et al., 2020; Cai et al., 2022; Qin et al., 2024). Recently, the authors in Tong et al. (2022); Dong et al. (2023b)

developed efficient algorithms for tensor RPCA, tensor completion and tensor regression under the Tucker decomposition using scaled gradient descent. In contrast to our work, the low-rank tensor in Tong et al. (2022); Dong et al. (2023b) is factorized as $(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}) \cdot \boldsymbol{S}$, and four factors are needed to be estimated, leading to a much more complicated nonconvex landscape than our case.

1.4 Paper Organization

The rest of this paper is organized as follows. Section 2 presents some notations and preliminaries. Section 3 describes the applications of the proposed ScaledGD method to the four problems studied in this paper with theoretical guarantees in terms of both statistical and computational complexities. In Section 4, we outline the proof for our main results. Section 5 illustrates the superior empirical performance of the proposed method. Finally, we conclude in Section 6. The proofs are deferred to the appendix.

2. Notations and Preliminaries

In this section, we introduce the notations and provide a concise overview of t-SVD, which establishes the groundwork for the development of our algorithms.

Notations. We use lowercase, bold lowercase, bold uppercase, and bold calligraphic letters for scalars, vectors, matrices, and tensors, respectively. The real and complex Euclidean spaces are denoted as $\mathbb{R}$ and $\mathbb{C}$, respectively. Superscript H and T denote conjugate transpose and transpose, respectively. For a three-way tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, we use $\boldsymbol{\mathcal{A}}(i,:,:)$, $\boldsymbol{\mathcal{A}}(:,i,:)$ and $\boldsymbol{\mathcal{A}}(:,:,i)$ to denote its i-th horizontal, lateral and frontal slice, respectively. The (i,j,k)-th entry of $\boldsymbol{\mathcal{A}}$ is denoted as $\boldsymbol{\mathcal{A}}_{i,j,k}$. For brevity, the frontal slice $\boldsymbol{\mathcal{A}}(:,:,i)$ and the lateral slice $\boldsymbol{\mathcal{A}}(:,i,:)$ are denoted compactly as $\boldsymbol{A}^{(i)}$ and $\boldsymbol{\mathcal{A}}_{(i)}$, respectively. The (i,j)-th mode-1, mode-2 and mode-3 tubes are denoted by $\boldsymbol{\mathcal{A}}(:,i,j)$, $\boldsymbol{\mathcal{A}}(i,:,j)$ and $\boldsymbol{\mathcal{A}}(i,j,:)$, respectively. For a matrix $\boldsymbol{A}$, its (i,j)-th entry is denoted as $\boldsymbol{A}_{i,j}$. The i-th row and the i-th column of $\boldsymbol{A}$ are denoted by $\boldsymbol{A}_{i,\cdot}$ and $\boldsymbol{A}_{\cdot,i}$, respectively. The $n \times n$ identity matrix is denoted by $\boldsymbol{I}_n$. The inner product between two matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{n_1 \times n_2}$ is defined as $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \operatorname{tr}(\boldsymbol{A}^H \boldsymbol{B})$, where $\operatorname{tr}(\cdot)$ denotes the matrix trace. The inner product between two tensors $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is defined as $\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle = \sum_{i=1}^{n_3} \langle \boldsymbol{A}^{(i)}, \boldsymbol{B}^{(i)} \rangle$. The facewise product between two tensors $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{n_2 \times n_4 \times n_3}$, denoted by $\boldsymbol{\mathcal{A}} \triangle \boldsymbol{\mathcal{B}}$, is a tensor $\boldsymbol{\mathcal{C}} \in \mathbb{C}^{n_1 \times n_4 \times n_3}$ such that each frontal slice of $\boldsymbol{\mathcal{C}}$ is the matrix multiplication of the corresponding frontal slices of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$, that is, $\boldsymbol{C}^{(i)} = \boldsymbol{A}^{(i)} \boldsymbol{B}^{(i)}$ (Kernfeld et al., 2015). Let $a \vee b = \max(a,b)$, $a \wedge b = \min(a,b)$, and $[a] := \{1, 2, \ldots, a\}$. 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. We use the terminology "with high probability" to denote the event happens with probability at least $1 - c_1 n^{-c_2}$, where c_1 and c_2 denote positive universal constants, which are allowed to differ from line to line.

Some norms of vector, matrix and tensor are used. The ℓ_1, ℓ_∞, $\ell_{2,\infty}$ and Frobenius norms of $\boldsymbol{\mathcal{A}}$ are defined as $\|\boldsymbol{\mathcal{A}}\|_1 = \sum_{i,j,k} |\boldsymbol{\mathcal{A}}_{i,j,k}|$, $\|\boldsymbol{\mathcal{A}}\|_\infty = \max_{i,j,k} |\boldsymbol{\mathcal{A}}_{i,j,k}|$, $\|\boldsymbol{\mathcal{A}}\|_{2,\infty} = \max_i \|\boldsymbol{\mathcal{A}}(i,:,:)\|_F$ and $\|\boldsymbol{\mathcal{A}}\|_F = \sqrt{\sum_{i,j,k} |\boldsymbol{\mathcal{A}}_{i,j,k}|^2}$, respectively. For $\boldsymbol{v} \in \mathbb{C}^n$, its ℓ_2 norm is denoted as $\|\boldsymbol{v}\|_2 = \sqrt{\sum_i |v_i|^2}$ and we use ℓ_0 norm to denote the number of nonzero elements

in $\boldsymbol{v}$. 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 $\boldsymbol{A}$. The matrix nuclear norm of $\boldsymbol{A}$ is $\|\boldsymbol{A}\|_* = \sum_i \sigma_i(\boldsymbol{A})$.

2.1 Tensor Singular Value Decomposition

The framework of tensor singular value decomposition (t-SVD) is based on the t-product under an invertible linear transform L (Kernfeld et al., 2015). In this paper, the transformation matrix $\boldsymbol{\Phi}$ defining the transform L is restricted to be orthogonal, i.e., $\boldsymbol{\Phi} \in \mathbb{C}^{n_3 \times n_3}$ satisfying

$$\boldsymbol{\Phi}\boldsymbol{\Phi}^H = \boldsymbol{\Phi}^H\boldsymbol{\Phi} = \ell \boldsymbol{I}_{n_3}, \tag{5}$$

where $\ell > 0$ is a constant. We define the associated linear transform $L(\cdot)$ with its inverse mapping $L^{-1}(\cdot)$ on any $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as

$$\bar{\boldsymbol{\mathcal{A}}} := L(\boldsymbol{\mathcal{A}}) = \boldsymbol{\mathcal{A}} \times_3 \boldsymbol{\Phi} \quad \text{and} \quad L^{-1}(\boldsymbol{\mathcal{A}}) := \boldsymbol{\mathcal{A}} \times_3 \boldsymbol{\Phi}^{-1}, \tag{6}$$

where $\times_3$ denotes the mode-3 tensor-matrix product (Kolda and Bader, 2009), i.e., for any $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ and $\boldsymbol{M} \in \mathbb{C}^{n_4 \times n_3}$, $\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{A}} \times_3 \boldsymbol{M} \in \mathbb{C}^{n_1 \times n_2 \times n_4}$ such that

$$\boldsymbol{\mathcal{B}}_{i,j,l} = \sum_{k=1}^{n_3} \boldsymbol{\mathcal{A}}_{i,j,k} \boldsymbol{M}_{l,k}, \quad i \in [n_1], j \in [n_2], l \in [n_4].$$

Definition 1 (t-product (Kernfeld et al., 2015)) *The t-product of any $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{n_2 \times n_4 \times n_3}$ under transform L in (6), is a unique tensor $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}} \in \mathbb{C}^{n_1 \times n_4 \times n_3}$ such that $L(\boldsymbol{\mathcal{C}}) = L(\boldsymbol{\mathcal{A}}) \triangle L(\boldsymbol{\mathcal{B}})$.*

We denote $\bar{\boldsymbol{A}} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3}$ as the block diagonal matrix with its i-th diagonal block corresponding to the i-th frontal slice $\bar{\boldsymbol{A}}^{(i)}$ of $\bar{\boldsymbol{\mathcal{A}}}$ in (6), i.e.,

$$\bar{\boldsymbol{A}} = \mathtt{bdiag}(\bar{\boldsymbol{\mathcal{A}}}) = \begin{bmatrix} \bar{\boldsymbol{A}}^{(1)} & & \\ & \ddots & \\ & & \bar{\boldsymbol{A}}^{(n_3)} \end{bmatrix}, \tag{7}$$

where $\mathtt{bdiag}$ is an operator that maps $\bar{\boldsymbol{\mathcal{A}}}$ to $\bar{\boldsymbol{A}}$. Using (5), we have the following properties:

$$\|\boldsymbol{\mathcal{A}}\|_F = \frac{1}{\sqrt{\ell}}\|\bar{\boldsymbol{\mathcal{A}}}\|_F = \frac{1}{\sqrt{\ell}}\|\bar{\boldsymbol{A}}\|_F \quad \text{and} \quad \langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle = \frac{1}{\ell}\langle \bar{\boldsymbol{\mathcal{A}}}, \bar{\boldsymbol{\mathcal{B}}} \rangle = \frac{1}{\ell}\langle \bar{\boldsymbol{A}}, \bar{\boldsymbol{B}} \rangle. \tag{8}$$

Definition 2 (Conjugate transpose (Song et al., 2020)) *The conjugate transpose of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ under L is the tensor $\boldsymbol{\mathcal{A}}^H \in \mathbb{C}^{n_2 \times n_1 \times n_3}$ satisfying $L(\boldsymbol{\mathcal{A}}^H)^{(i)} = (L(\boldsymbol{\mathcal{A}})^{(i)})^H$, $i = 1, 2, \ldots, n_3$.*

Definition 3 (Identity tensor (Song et al., 2020)) *The identity tensor $\boldsymbol{\mathcal{I}}_n \in \mathbb{C}^{n \times n \times n_3}$ under L is a tensor such that each frontal slice of $L(\boldsymbol{\mathcal{I}}_n) = \bar{\boldsymbol{\mathcal{I}}}_n$ is the identity matrix $\boldsymbol{I}_n$. Then $\boldsymbol{\mathcal{I}}_n = L^{-1}(\bar{\boldsymbol{\mathcal{I}}}_n)$ gives the identity tensor under L.*

Definition 4 (Orthogonal tensor (Song et al., 2020)) *A tensor $\boldsymbol{\mathcal{Q}} \in \mathbb{C}^{n \times n \times n_3}$ is orthogonal if it satisfies $\boldsymbol{\mathcal{Q}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}^H = \boldsymbol{\mathcal{Q}}^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{I}}_n$.*

Definition 5 (Invertible tensor) *A tensor $\mathcal{A} \in \mathbb{R}^{r \times r \times n_3}$ is invertible if there exists a tensor $\mathcal{B} \in \mathbb{R}^{r \times r \times n_3}$ such that $\mathcal{A} *_{\boldsymbol{\Phi}} \mathcal{B} = \mathcal{B} *_{\boldsymbol{\Phi}} \mathcal{A} = \mathcal{I}_r$. The set of invertible tensors in $\mathbb{R}^{r \times r \times n_3}$ is denoted by $\mathrm{GL}(r)$.*

Definition 6 (t-SVD (Song et al., 2020)) *Let L be any invertible linear transform in (6). For any $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, it can be factorized as $\mathcal{A} = \mathcal{U} *_{\boldsymbol{\Phi}} \mathcal{G} *_{\boldsymbol{\Phi}} \mathcal{V}^H$, where $\mathcal{U} \in \mathbb{C}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{C}^{n_2 \times n_2 \times n_3}$ are orthogonal, and $\mathcal{G} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor, whose frontal slices are diagonal matrices.*

For any $\mathcal{A}$, we have the following relationship between its t-SVD and the matrix SVD of its block-diagonal matrix (Kernfeld et al., 2015):

$$\mathcal{A} = \mathcal{U} *_{\boldsymbol{\Phi}} \mathcal{G} *_{\boldsymbol{\Phi}} \mathcal{V}^H \Leftrightarrow \bar{A} = \bar{U}\bar{G}\bar{V}^H.$$

In words, the t-product in the spatial domain corresponds to matrix multiplication of the frontal slices in the transformed domain. Note that when $n_3 = 1$, the operator $*_{\boldsymbol{\Phi}}$ reduces to matrix multiplication.

Definition 7 (Tensor multi-rank and tubal rank (Kilmer et al., 2021)) *The multi-rank of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is a vector $\boldsymbol{r} \in \mathbb{R}^{n_3}$, with its i-th entry being the rank of the i-th frontal slice of $\bar{\mathcal{A}}$, i.e., $r_i = \mathrm{rank}(\bar{A}^{(i)})$. Let $\mathcal{A} = \mathcal{U} *_{\boldsymbol{\Phi}} \mathcal{G} *_{\boldsymbol{\Phi}} \mathcal{V}^H$ be the t-SVD of $\mathcal{A}$. The tensor tubal rank $\mathrm{rank}_t(\mathcal{A})$ under L is defined as the number of nonzero singular tubes of $\mathcal{G}$, i.e.,*

$$\mathrm{rank}_t(\mathcal{A}) = \sharp\{i : \mathcal{G}(i, i, :) \neq \mathbf{0}\} = \max(r_1, \ldots, r_{n_3}).$$

We denote

$$\mathbb{P}_r(\mathcal{A}) = \underset{\widetilde{\mathcal{A}}:\mathrm{rank}_t(\widetilde{\mathcal{A}}) \leq r}{\arg\min} \|\mathcal{A} - \widetilde{\mathcal{A}}\|_F^2 \tag{9}$$

as the tubal rank-r approximation of $\mathcal{A}$, which is given by the top-r t-SVD of $\mathcal{A}$ by the tensor Eckart-Young theorem (Zhang et al., 2014; Kilmer et al., 2021).

Definition 8 (Tensor nuclear norm and spectral norm (Lu et al., 2019)) *The tensor nuclear norm of $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ under L is defined as $\|\mathcal{A}\|_* = \frac{1}{\ell} \sum_{i=1}^{n_3} \|\bar{A}^{(i)}\|_* = \frac{1}{\ell}\|\bar{A}\|_*$. The spectral norm of $\mathcal{A}$ is defined as $\|\mathcal{A}\| = \|\bar{A}\|$.*

3. Main Results

This section is devoted to introducing ScaledGD and establishing its performance guarantee for various low-rank tensor estimation problems.

3.1 Models and Assumptions

Suppose that the ground truth tubal rank-r tensor $\mathcal{X}_\star$ with multi-rank $\boldsymbol{r}$ admits the following compact t-SVD $\mathcal{X}_\star = \mathcal{U}_\star *_{\boldsymbol{\Phi}} \mathcal{G}_\star *_{\boldsymbol{\Phi}} \mathcal{V}_\star^H$, where $\mathcal{U}_\star \in \mathbb{R}^{n_1 \times r \times n_3}$, $\mathcal{V}_\star \in \mathbb{R}^{n_2 \times r \times n_3}$, and $\mathcal{G}_\star \in \mathbb{R}^{r \times r \times n_3}$. Define the ground truth low-rank factors as

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

8

so that $\boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{L}}_\star *_\Phi \boldsymbol{\mathcal{R}}_\star^H$. Here, the "square root" of a tensor $\boldsymbol{\mathcal{A}}$, denoted by $\boldsymbol{\mathcal{A}}^{\frac{1}{2}}$, is obtained by setting $\boldsymbol{\mathcal{A}}^{\frac{1}{2}} := \boldsymbol{\mathcal{B}} = L^{-1}(\bar{\boldsymbol{\mathcal{B}}})$, where the i-th frontal slice of $\bar{\boldsymbol{\mathcal{B}}}$ is $\bar{\boldsymbol{B}}^{(i)} = (\bar{\boldsymbol{A}}^{(i)})^{\frac{1}{2}}$. The factored representation is not unique in that for any invertible tensor $\boldsymbol{\mathcal{Q}} \in \mathrm{GL}(r)$, one has $\boldsymbol{\mathcal{L}}_\star *_\Phi \boldsymbol{\mathcal{R}}_\star^H = (\boldsymbol{\mathcal{L}}_\star *_\Phi \boldsymbol{\mathcal{Q}}) *_\Phi (\boldsymbol{\mathcal{R}}_\star *_\Phi \boldsymbol{\mathcal{Q}}^{-H})^H$. We define the following two important singular values of tensor $\boldsymbol{\mathcal{X}}_\star$ as

$$\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star) = \max\{\bar{\boldsymbol{\mathcal{G}}}_{i,i,k}|\bar{\boldsymbol{\mathcal{G}}}_{i,i,k} > 0, i \leq r_k, k \in [n_3]\}$$

$$\text{and} \quad \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) = \min\{\bar{\boldsymbol{\mathcal{G}}}_{i,i,k}|\bar{\boldsymbol{\mathcal{G}}}_{i,i,k} > 0, i \leq r_k, k \in [n_3]\}.$$

We first introduce the condition number of the tensor $\boldsymbol{\mathcal{X}}_\star$.

Definition 9 (Condition number) *The condition number κ of $\boldsymbol{\mathcal{X}}_\star$ is defined as*

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

Another parameter is the incoherence parameter, which is crucial in determining the well-posedness of low-rank tensor estimation.

Definition 10 (Incoherence) *For a tubal rank-r tensor $\boldsymbol{\mathcal{X}}_\star \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, assume that the multi-rank of $\boldsymbol{\mathcal{X}}_\star$ is $\boldsymbol{r}$ and it has the t-SVD $\boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{U}}_\star *_\Phi \boldsymbol{\mathcal{G}}_\star *_\Phi \boldsymbol{\mathcal{V}}_\star^H$. Then $\boldsymbol{\mathcal{X}}_\star$ is said to satisfy the tensor incoherence conditions with parameter μ if*

$$\|\boldsymbol{\mathcal{U}}_\star\|_{2,\infty} = \max_{i\in[n_1]} \|\boldsymbol{\mathcal{U}}_\star^H *_\Phi \mathring{\mathfrak{e}}_i\|_F \leq \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \quad \text{and} \quad \|\boldsymbol{\mathcal{V}}_\star\|_{2,\infty} = \max_{j\in[n_2]} \|\boldsymbol{\mathcal{V}}_\star^H *_\Phi \mathring{\mathfrak{e}}_j\|_F \leq \sqrt{\frac{\mu s_r}{n_2 n_3 \ell}}. \tag{10}$$

Here, $\mathring{\mathfrak{e}}_i$ is a tensor of size $n_1 \times 1 \times n_3$ with the entries of the $(i,1)$-th mode-3 tube of $L(\mathring{\mathfrak{e}}_i)$ equaling 1 and the rest equaling 0, and $s_r = \sum_{k=1}^{n_3} r_k$.

Notice that when all the r_k's are equal to the tubal rank r, i.e., $s_r = n_3 r$, the above conditions will be equivalent to the ones in Lu (2021, Proposition 1). Roughly speaking, a small incoherence parameter ensures that the tensor columns are not correlated with the standard tensor basis, i.e., the energy of the tensor is evenly distributed across its entries. To track the performance of ScaledGD iterates $\boldsymbol{\mathcal{F}}_t = [\boldsymbol{\mathcal{L}}_t^H, \boldsymbol{\mathcal{R}}_t^H]^H$, one needs a distance metric that properly takes account of the factor ambiguity due to invertible transforms. Motivated by the analysis in Tong et al. (2021a), we consider the following distance metric that resolves the ambiguity in the t-SVD.

Definition 11 (Distance metric) *Let $\boldsymbol{\mathcal{F}} = \begin{bmatrix} \boldsymbol{\mathcal{L}} \\ \boldsymbol{\mathcal{R}} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times r \times n_3}$ and $\boldsymbol{\mathcal{F}}_\star = \begin{bmatrix} \boldsymbol{\mathcal{L}}_\star \\ \boldsymbol{\mathcal{R}}_\star \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times r \times n_3}$, denote*

$$\mathrm{dist}(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}_\star) = \sqrt{\inf_{\boldsymbol{\mathcal{Q}}\in\mathrm{GL}(r)} \|(\boldsymbol{\mathcal{L}} *_\Phi \boldsymbol{\mathcal{Q}} - \boldsymbol{\mathcal{L}}_\star) *_\Phi \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \|(\boldsymbol{\mathcal{R}} *_\Phi \boldsymbol{\mathcal{Q}}^{-H} - \boldsymbol{\mathcal{R}}_\star) *_\Phi \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2}. \tag{11}$$

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

As indicated in Appendix A, for the ScaledGD iterates $\{\boldsymbol{\mathcal{F}}_t\}$, the optimal alignment tensors $\{\boldsymbol{\mathcal{Q}}_t\}$ always exist and hence are well-defined.

3.2 Tensor RPCA

Suppose we have observed a data tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of the form $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}_\star + \boldsymbol{\mathcal{S}}_\star$, where $\boldsymbol{\mathcal{X}}_\star$ is a low-rank tensor with tubal rank-r and $\boldsymbol{\mathcal{S}}_\star$ is a sparse tensor—in which the number of nonzero entries is much smaller than its ambient dimension—modeling corruptions in the observations due to sensor failures, malicious attacks, or other system errors. Given $\boldsymbol{\mathcal{Y}}$ and r, the goal of tensor RPCA is to reliably estimate the two tensors $\boldsymbol{\mathcal{X}}_\star$ and $\boldsymbol{\mathcal{S}}_\star$.

Following the matrix case in Chandrasekaran et al. (2011); Netrapalli et al. (2014), we consider a deterministic sparsity model for $\boldsymbol{\mathcal{S}}_\star$, in which $\boldsymbol{\mathcal{S}}_\star$ contains at most α-fraction of nonzero entries per tube.

Definition 12 *A sparse tensor* $\boldsymbol{\mathcal{S}}_\star \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *is* α-*sparse, i.e.,* $\boldsymbol{\mathcal{S}}_\star \in \mathscr{S}_\alpha$, *where we denote*

$$\mathscr{S}_\alpha := \{\boldsymbol{\mathcal{S}} \in \mathbb{R}^{n_1 \times n_2 \times n_3} : \|\boldsymbol{\mathcal{S}}(:,j,k)\|_0 \le \alpha n_1, \|\boldsymbol{\mathcal{S}}(i,:,k)\|_0 \le \alpha n_2, \|\boldsymbol{\mathcal{S}}(i,j,:)\|_0 \le \alpha n_3,$$
$$\text{for } i \in [n_1], j \in [n_2], k \in [n_3]\}.$$

Motivated by the works in Tong et al. (2021a); Cai et al. (2021), we parameterize $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}^H$ by two low-rank factors $\boldsymbol{\mathcal{L}} \in \mathbb{R}^{n_1 \times r \times n_3}$ and $\boldsymbol{\mathcal{R}} \in \mathbb{R}^{n_2 \times r \times n_3}$ that are more memory-efficient, and instead optimize over the factors by solving the following optimization problem:

$$\min_{\boldsymbol{\mathcal{F}} = [\boldsymbol{\mathcal{L}}^H, \boldsymbol{\mathcal{R}}^H]^H \in \mathbb{R}^{(n_1+n_2) \times r \times n_3}, \boldsymbol{\mathcal{S}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} f(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{S}}) := \frac{1}{2} \|\boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}^H + \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{Y}}\|_F^2. \tag{12}$$

Despite the nonconvexity of the objective function, one might be tempted to update the tensor factors via gradient descent, which, however, likely to converge slowly even for moderately ill-conditioned tensors (Yi et al., 2016; Tong et al., 2021a; Han et al., 2022). Similar to the algorithms in Tong et al. (2021a); Dong et al. (2023b), our algorithm alternates between corruption removal and factor refinements, where $(\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{R}})$ are updated via the proposed ScaledGD algorithm, and $\boldsymbol{\mathcal{S}}$ is updated by soft thresholding. Specifically, we use the following soft-shrinkage operator $\mathcal{T}_\zeta(\cdot)$ that sets entries with magnitudes smaller than ζ to 0, while uniformly shrinking the magnitudes of the other entries by ζ, defined as

$$[\mathcal{T}_\zeta(\boldsymbol{\mathcal{M}})]_{i,j,k} := \text{sgn}([\boldsymbol{\mathcal{M}}]_{i,j,k}) \cdot \max(0, |[\boldsymbol{\mathcal{M}}]_{i,j,k}| - \zeta). \tag{13}$$

The sparse outlier tensor is updated via

$$\boldsymbol{\mathcal{S}}_{t+1} = \mathcal{T}_{\zeta_{t+1}}(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H) \tag{14}$$

with the schedule ζ_t to be specified shortly.

To complete the algorithm description, we still need to specify how to initialize the algorithm. We start with initializing the sparse tensor by $\boldsymbol{\mathcal{S}}_0 = \mathcal{T}_{\zeta_0}(\boldsymbol{\mathcal{Y}})$ to remove the obvious outliers. Next, for the low-rank component, we set $\boldsymbol{\mathcal{L}}_0 = \boldsymbol{\mathcal{U}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0^{\frac{1}{2}}$ and $\boldsymbol{\mathcal{R}}_0 = \boldsymbol{\mathcal{V}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0^{\frac{1}{2}}$, where $\boldsymbol{\mathcal{U}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}_0^H$ is the best tubal rank-r approximation of $\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{S}}_0$. Combining all the steps mentioned above, we can now formally present the algorithm in Algorithm 1, which we dub ScaledGD for simplicity. ScaledGD costs only $\mathcal{O}(n_1 n_2 n_3 r + n_1 n_2 n_3^2 + (n_1 + n_2) n_3 r^2 + n_3 r^3)$ flops per iteration provided $r \ll n_1 \wedge n_2$. For some special transforms, e.g., DFT, the per-iteration complexity is $\mathcal{O}(n_1 n_2 n_3 r + n_1 n_2 n_3 \log(n_3) + (n_1 + n_2) n_3 r^2 + n_3 r^3)$. The breakdown of ScaledGD's complexity is provided in Appendix G.

Algorithm 1 ScaledGD for tensor robust PCA with spectral initialization

---

Input: Observed tensor $\boldsymbol{\mathcal{Y}}$, the transformation matrix $\boldsymbol{\Phi}$ associated with the transform L, the tubal rank r, learning rate η, maximum number of iterations T, and threshold schedule $\{\zeta_t\}_{t=0}^T$.

    Spectral initialization: Let $\boldsymbol{\mathcal{U}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}_0^H$ be the top-r t-SVD of $\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{S}}_0$, where $\boldsymbol{\mathcal{S}}_0 = \mathcal{T}_{\zeta_0}(\boldsymbol{\mathcal{Y}})$.

$$\text{Set} \quad \boldsymbol{\mathcal{L}}_0 = \boldsymbol{\mathcal{U}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0^{\frac{1}{2}} \quad \text{and} \quad \boldsymbol{\mathcal{R}}_0 = \boldsymbol{\mathcal{V}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0^{\frac{1}{2}}.$$

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

$$\boldsymbol{\mathcal{S}}_{t+1} = \mathcal{T}_{\zeta_{t+1}}(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H)$$

$$\boldsymbol{\mathcal{L}}_{t+1} = \boldsymbol{\mathcal{L}}_t - \eta(\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H + \boldsymbol{\mathcal{S}}_{t+1} - \boldsymbol{\mathcal{Y}}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_t^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t)^{-1} \quad (15)$$

$$\boldsymbol{\mathcal{R}}_{t+1} = \boldsymbol{\mathcal{R}}_t - \eta(\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H + \boldsymbol{\mathcal{S}}_{t+1} - \boldsymbol{\mathcal{Y}})^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_t^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_t)^{-1}.$$

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

Theoretical guarantees. The following theorem establishes that ScaledGD algorithm—with proper choices of the tuning parameters, recovers the ground truth tensor $\boldsymbol{\mathcal{X}}_\star$, as long as the fraction of corruptions is not too large. For convenience, we denote $n_{(1)} = \max(n_1, n_2)$ and $n_{(2)} = \min(n_1, n_2)$ in the following.

Theorem 13 *Let $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}_\star + \boldsymbol{\mathcal{S}}_\star \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where $\boldsymbol{\mathcal{X}}_\star$ is a multi-rank $\boldsymbol{r}$ tensor with μ-incoherence and $\boldsymbol{\mathcal{S}}_\star$ is α-sparse. Suppose that the thresholding values $\{\zeta_t\}_{t=0}^\infty$ obey that $\|\boldsymbol{\mathcal{X}}_\star\|_\infty \leq \zeta_0 \leq 2\|\boldsymbol{\mathcal{X}}_\star\|_\infty$ and $\zeta_{t+1} = \rho\zeta_t, t \geq 1$, for some properly tuned $\zeta_1 = 3\frac{\mu s_r}{n_3\sqrt{n_1 n_2 \ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$ and $\frac{1}{4} \leq \eta \leq \frac{8}{9}$, where $\rho = 1 - 0.6\eta$. Then the iterates of ScaledGD with spectral initialization satisfy*

$$\|\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H - \boldsymbol{\mathcal{X}}_\star\|_F \leq \frac{0.03}{\sqrt{\ell}}\rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star),$$

$$\|\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H - \boldsymbol{\mathcal{X}}_\star\|_\infty \leq 3\rho^t \frac{\mu s_r}{n_3\sqrt{n_1 n_2 \ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star),$$

$$\|\boldsymbol{\mathcal{S}}_t - \boldsymbol{\mathcal{S}}_\star\|_\infty \leq 6\rho^{t-1}\frac{\mu s_r}{n_3\sqrt{n_1 n_2 \ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) \quad \text{and} \quad \text{supp}(\boldsymbol{\mathcal{S}}_t) \subseteq \text{supp}(\boldsymbol{\mathcal{S}}_\star) \quad (16)$$

for all $t \geq 1$, as long as the level of corruptions obeys $\alpha \leq \frac{n_3\sqrt{n_{(2)}n_2}}{10^4 \mu s_r^{1.5}\kappa(n_1 + n_2 n_3)} \wedge \frac{n_3}{10^5 \mu^2 s_r^2 \kappa^2}$.

    Note that the value of ρ was used to simplify the proof, and it should not be considered as an optimal convergence rate. Theorem 13 implies that upon appropriate choices of parameters, if the level of corruptions satisfies $\alpha \lesssim \frac{n_3\sqrt{n_{(2)}n_2}}{\mu s_r^{1.5}\kappa(n_1 + n_2 n_3)} \wedge \frac{n_3}{\mu^2 s_r^2 \kappa^2}$, we can ensure that the proposed ScaledGD algorithm—starting from a carefully designed spectral initialization, converges linearly to the ground truth tensor $\boldsymbol{\mathcal{X}}_\star$ in both the Frobenius norm and the entrywise ℓ_∞ norm at a constant rate, which is independent of the condition number, even when the gross corruptions are arbitrary. Note that the choice of parameters relies

11

on the knowledge of $\mathcal{X}_\star$, which is usually unknown in practice. Thus, Theorem 13 can be considered as a proof for the existence of the appropriate parameters.

The proof of our convergence theorem follows the route established in Cai et al. (2021). However, the algorithm in Cai et al. (2021) is designed for matrices, while the extension from matrices to tensors is not trivial as different mathematical tools are required. For example, we need to use the property of t-product to prove some bounds on norms of $\mathcal{S}$, and we have to interconvert between the original and transformed domains in the proofs.

3.3 Tensor Completion

We now consider the tensor completion problem. Let $\mathcal{X}_\star \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an unknown tensor and it has tubal rank $\mathrm{rank}_t(\mathcal{X}_\star) = r$. We assume to observe the entries of $\mathcal{X}_\star$ at locations given by a set $\boldsymbol{\Omega} = \{(i, j, k) | \delta_{ijk} = 1\}$, where δ_{ijk}'s are independent and identically distributed (i.i.d.) Bernoulli variables taking value one with probability p and zero with probability $1 - p$. We denote such a Bernoulli sampling by $\boldsymbol{\Omega} \sim \mathrm{Ber}(p)$. The goal of tensor completion is to recover the tensor $\mathcal{X}_\star$ from its partial observation $\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\mathcal{X}_\star)$, where $\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a projection such that

$$[\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\mathcal{X}_\star)]_{i,j,k} = \begin{cases} [\mathcal{X}_\star]_{i,j,k}, & \text{if } (i, j, k) \in \boldsymbol{\Omega}, \\ 0, & \text{otherwise.} \end{cases}$$

This can be achieved by minimizing the loss function

$$\min_{\boldsymbol{\mathcal{F}} = [\boldsymbol{\mathcal{L}}^H, \boldsymbol{\mathcal{R}}^H]^H \in \mathbb{R}^{(n_1+n_2) \times r \times n_3}} f(\boldsymbol{\mathcal{F}}) := \frac{1}{2p} \| \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}^H - \mathcal{X}_\star) \|_F^2. \tag{17}$$

Similar to tensor RPCA, the underlying low-rank tensor $\mathcal{X}_\star$ is required to be incoherent (cf. Definition 10) to avoid ill-posedness. One typical strategy to ensure the incoherence condition in the matrix setting is to trim the rows of the factors after the gradient update (Chen and Wainwright, 2015). However, we need to be careful here because we need to preserve the equivariance with respect to invertible transforms. Again, inspired by Tong et al. (2021a), we introduce the following projection operator: for every $\widetilde{\boldsymbol{\mathcal{F}}} = \begin{bmatrix} \widetilde{\boldsymbol{\mathcal{L}}} \\ \widetilde{\boldsymbol{\mathcal{R}}} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2) \times r \times n_3}$,

$$\mathcal{P}_\varsigma(\widetilde{\boldsymbol{\mathcal{F}}}) := \underset{\boldsymbol{\mathcal{F}} \in \mathbb{R}^{(n_1+n_2) \times r \times n_3}}{\arg\min} \| (\boldsymbol{\mathcal{L}} - \widetilde{\boldsymbol{\mathcal{L}}}) *_{\boldsymbol{\Phi}} (\widetilde{\boldsymbol{\mathcal{R}}}^H *_{\boldsymbol{\Phi}} \widetilde{\boldsymbol{\mathcal{R}}})^{\frac{1}{2}} \|_F^2 + \| (\boldsymbol{\mathcal{R}} - \widetilde{\boldsymbol{\mathcal{R}}}) *_{\boldsymbol{\Phi}} (\widetilde{\boldsymbol{\mathcal{L}}}^H *_{\boldsymbol{\Phi}} \widetilde{\boldsymbol{\mathcal{L}}})^{\frac{1}{2}} \|_F^2$$

$$\text{s.t.} \quad \sqrt{n_1} \| \boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} (\widetilde{\boldsymbol{\mathcal{R}}}^H *_{\boldsymbol{\Phi}} \widetilde{\boldsymbol{\mathcal{R}}})^{\frac{1}{2}} \|_{2,\infty} \vee \sqrt{n_2} \| \boldsymbol{\mathcal{R}} *_{\boldsymbol{\Phi}} (\widetilde{\boldsymbol{\mathcal{L}}}^H *_{\boldsymbol{\Phi}} \widetilde{\boldsymbol{\mathcal{L}}})^{\frac{1}{2}} \|_{2,\infty} \leq \varsigma, \tag{18}$$

where $\varsigma > 0$ is the projection radius. Fortunately, following the proof in Tong et al. (2021a, Proposition 7), this problem has a closed-form solution, as stated below.

Proposition 14 *The solution to* (18) *is given by* $\mathcal{P}_\varsigma(\widetilde{\boldsymbol{\mathcal{F}}}) = \begin{bmatrix} \boldsymbol{\mathcal{L}} \\ \boldsymbol{\mathcal{R}} \end{bmatrix}$, *where*

$$\boldsymbol{\mathcal{L}}(i, :, :) = \left(1 \wedge \frac{\varsigma}{\sqrt{n_1} \| \widetilde{\boldsymbol{\mathcal{L}}}(i, :, :) *_{\boldsymbol{\Phi}} \widetilde{\boldsymbol{\mathcal{R}}}^H \|_F} \right) \widetilde{\boldsymbol{\mathcal{L}}}(i, :, :), \quad i \in [n_1],$$

$$\text{and} \quad \boldsymbol{\mathcal{R}}(j, :, :) = \left(1 \wedge \frac{\varsigma}{\sqrt{n_2} \| \widetilde{\boldsymbol{\mathcal{R}}}(j, :, :) *_{\boldsymbol{\Phi}} \widetilde{\boldsymbol{\mathcal{L}}}^H \|_F} \right) \widetilde{\boldsymbol{\mathcal{R}}}(j, :, :), \quad j \in [n_2]. \tag{19}$$

Algorithm 2 ScaledGD for tensor completion with spectral initialization

---

Input: Partially observed data tensor $\mathcal{P}_\Omega(\mathcal{X}_\star)$, the transformation matrix $\mathbf{\Phi}$ associated with the transform L, the tubal rank r, learning rate η, and maximum number of iterations T.

Spectral initialization: Let $\mathcal{U}_0 *_\mathbf{\Phi} \mathcal{G}_0 *_\mathbf{\Phi} \mathcal{V}_0^H$ be the top-r t-SVD of $\frac{1}{p}\mathcal{P}_\Omega(\mathcal{X}_\star)$, and set

$$\begin{bmatrix} \mathcal{L}_0 \\ \mathcal{R}_0 \end{bmatrix} = \mathcal{P}_\varsigma\left(\begin{bmatrix} \mathcal{U}_0 *_\mathbf{\Phi} \mathcal{G}_0^{\frac{1}{2}} \\ \mathcal{V}_0 *_\mathbf{\Phi} \mathcal{G}_0^{\frac{1}{2}} \end{bmatrix} \right). \tag{20}$$

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

$$\begin{bmatrix} \mathcal{L}_{t+1} \\ \mathcal{R}_{t+1} \end{bmatrix} = \mathcal{P}_\varsigma\left(\begin{bmatrix} \mathcal{L}_t - \frac{\eta}{p}\mathcal{P}_\Omega(\mathcal{L}_t *_\mathbf{\Phi} \mathcal{R}_t^H - \mathcal{X}_\star) *_\mathbf{\Phi} \mathcal{R}_t *_\mathbf{\Phi} (\mathcal{R}_t^H *_\mathbf{\Phi} \mathcal{R}_t)^{-1} \\ \mathcal{R}_t - \frac{\eta}{p}\mathcal{P}_\Omega(\mathcal{L}_t *_\mathbf{\Phi} \mathcal{R}_t^H - \mathcal{X}_\star)^H *_\mathbf{\Phi} \mathcal{L}_t *_\mathbf{\Phi} (\mathcal{L}_t^H *_\mathbf{\Phi} \mathcal{L}_t)^{-1} \end{bmatrix} \right). \tag{21}$$

Output: The recovered low-rank tensor $\mathcal{X}_T = \mathcal{L}_T *_\mathbf{\Phi} \mathcal{R}_T^H$.

With this projection operator in hand, we are ready to propose our ScaledGD method with the spectral initialization for solving tensor completion, as described in Algorithm 2. The computational cost at each iteration is $\mathcal{O}(n_1 n_2 n_3 r + p n_1 n_2 n_3^2 + (n_1 + n_2)n_3 r^2 + n_3 r^3)$ for general linear transforms, with reduction to $\mathcal{O}(n_1 n_2 n_3 r + p n_1 n_2 n_3 \log(n_3) + (n_1 + n_2)n_3 r^2 + n_3 r^3)$ for DFT, which is much lower compared to the algorithm in Lu et al. (2019).

Theoretical guarantees. Encouragingly, we can guarantee that ScaledGD provably recovers the ground truth tensor, as long as the sample size is sufficiently large.

Theorem 15 *Suppose that $\mathcal{X}_\star$ is μ-incoherent, and that p satisfies $p \geq c\left(\frac{\mu s_r(n_1+n_2)\log((n_1 \vee n_2)n_3)}{n_1 n_2 n_3} \vee \frac{\mu^2 s_r^2 \kappa^4 \ell \log((n_1 \vee n_2)n_3)}{(n_1 \wedge n_2)n_3^2}\right)$ for some sufficiently large constant c. Set the projection radius as $\varsigma \geq c_\varsigma \sqrt{\frac{\mu s_r}{n_3 \ell}}\bar{\sigma}_1(\mathcal{X}_\star)$ for some constant $c_\varsigma \geq 1.02$. If the step size obeys $0 < \eta \leq \frac{2}{3}$, then with high probability, for all $t \geq 0$, the iterates of ScaledGD in (21) satisfy*

$$\mathrm{dist}(\mathcal{F}_t, \mathcal{F}_\star) \leq (1 - 0.6\eta)^t 0.02\bar{\sigma}_{s_r}(\mathcal{X}_\star) \quad \text{and} \quad \|\mathcal{L}_t *_\mathbf{\Phi} \mathcal{R}_t^H - \mathcal{X}_\star\|_F \leq (1 - 0.6\eta)^t 0.03\bar{\sigma}_{s_r}(\mathcal{X}_\star).$$

Theorem 15 establishes that the distance $\mathrm{dist}(\mathcal{F}_t, \mathcal{F}_\star)$ contracts linearly at a constant rate, as long as the probability of observation satisfies $p \gtrsim \left(\frac{\mu s_r(n_1+n_2)\log((n_1 \vee n_2)n_3)}{n_1 n_2 n_3} \vee \frac{\mu^2 s_r^2 \kappa^4 \ell \log((n_1 \vee n_2)n_3)}{(n_1 \wedge n_2)n_3^2}\right)$. To reach an ϵ-accurate estimate, i.e., $\|\mathcal{L}_t *_\mathbf{\Phi} \mathcal{R}_t^H - \mathcal{X}_\star\|_F \leq \epsilon\bar{\sigma}_{s_r}(\mathcal{X}_\star)$, ScaledGD takes at most $\mathcal{O}(\log(1/\epsilon))$ iterations, which is independent of the condition number, as long as the sample complexity is large enough.

3.4 Robust Tensor Completion

Assume that we partially observe some entries of a data tensor $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of the form $\mathcal{Y} = \mathcal{X}_\star + \mathcal{S}_\star$, where $\mathcal{X}_\star$ is a tubal rank-r tensor and $\mathcal{S}_\star$ is a sparse tensor. The task of robust tensor completion is to recover the two tensors $\mathcal{X}_\star$ and $\mathcal{S}_\star$ from the observations

---

Algorithm 3 ScaledGD for robust tensor completion with spectral initialization

---

Input: Partially observed data tensor $\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{Y})$, the transformation matrix $\boldsymbol{\Phi}$ associated with the transform L, the tubal rank r, learning rate η, maximum number of iterations T, and threshold schedule $\{\zeta_t\}_{t=0}^T$.

    Spectral initialization: Let $\mathcal{U}_0 *_{\boldsymbol{\Phi}} \mathcal{G}_0 *_{\boldsymbol{\Phi}} \mathcal{V}_0^H$ be the top-r t-SVD of $\frac{1}{p}\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{Y} - \mathcal{S}_0)$, where $\mathcal{S}_0 = \mathcal{T}_{\zeta_0}(\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{Y}))$.

$$\text{Set} \quad \mathcal{L}_0 = \mathcal{U}_0 *_{\boldsymbol{\Phi}} \mathcal{G}_0^{\frac{1}{2}} \quad \text{and} \quad \mathcal{R}_0 = \mathcal{V}_0 *_{\boldsymbol{\Phi}} \mathcal{G}_0^{\frac{1}{2}}.$$

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

$$
\begin{aligned}
\mathcal{S}_{t+1} &= \mathcal{T}_{\zeta_{t+1}}(\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{Y} - \mathcal{L}_t *_{\boldsymbol{\Phi}} \mathcal{R}_t^H)) \\
\mathcal{L}_{t+1} &= \mathcal{L}_t - \frac{\eta}{p}\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{L}_t *_{\boldsymbol{\Phi}} \mathcal{R}_t^H + \mathcal{S}_{t+1} - \mathcal{Y}) *_{\boldsymbol{\Phi}} \mathcal{R}_t *_{\boldsymbol{\Phi}} (\mathcal{R}_t^H *_{\boldsymbol{\Phi}} \mathcal{R}_t)^{-1} \quad (23) \\
\mathcal{R}_{t+1} &= \mathcal{R}_t - \frac{\eta}{p}\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{L}_t *_{\boldsymbol{\Phi}} \mathcal{R}_t^H + \mathcal{S}_{t+1} - \mathcal{Y})^H *_{\boldsymbol{\Phi}} \mathcal{L}_t *_{\boldsymbol{\Phi}} (\mathcal{L}_t^H *_{\boldsymbol{\Phi}} \mathcal{L}_t)^{-1}.
\end{aligned}
$$

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

$\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{Y}) = \mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{X}_\star + \mathcal{S}_\star)$, where the set of observed locations in $\boldsymbol{\Omega}$ is sampled independently according to the Bernoulli model with probability p. We also assume that $\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{S}_\star)$ is αp-sparse, i.e., it has at most αp-fraction of nonzero entries per tube for each mode.

    To avoid the nonconvex low-tubal-rank constraint on $\mathcal{X}$, we rewrite $\mathcal{X} = \mathcal{L} *_{\boldsymbol{\Phi}} \mathcal{R}^H$ and consider the following objective function:

$$\min_{\mathcal{F}=[\mathcal{L}^H, \mathcal{R}^H]^H \in \mathbb{R}^{(n_1+n_2) \times r \times n_3}, \mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} f(\mathcal{F}, \mathcal{S}) \coloneqq \frac{1}{2p}\|\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{L} *_{\boldsymbol{\Phi}} \mathcal{R}^H + \mathcal{S} - \mathcal{Y})\|_F^2. \quad (22)$$

Similar to Algorithm 1, we first initialize the sparse tensor by $\mathcal{S}_0 = \mathcal{T}_{\zeta_0}(\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{Y}))$ and take $\mathcal{L}_0 = \mathcal{U}_0 *_{\boldsymbol{\Phi}} \mathcal{G}_0^{\frac{1}{2}}$ and $\mathcal{R}_0 = \mathcal{V}_0 *_{\boldsymbol{\Phi}} \mathcal{G}_0^{\frac{1}{2}}$, where $\mathcal{U}_0 *_{\boldsymbol{\Phi}} \mathcal{G}_0 *_{\boldsymbol{\Phi}} \mathcal{V}_0^H$ is the best tubal rank-r approximation of $\frac{1}{p}\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{Y} - \mathcal{S}_0)$. Note that p^{-1} is necessary here to reweight the expectation of sampling operator (Recht, 2011). In the phase of iterative updates, we alternatively update the sparse and low-rank components in the fashion of ScaledGD, formally stated in Algorithm 3. For any invertible linear transforms, the per-iteration complexity is $\mathcal{O}(n_1 n_2 n_3 r + p n_1 n_2 n_3^2 + (n_1 + n_2)n_3 r^2 + n_3 r^3)$. For some special transforms, e.g., DFT, the per-iteration complexity is $\mathcal{O}(n_1 n_2 n_3 r + p n_1 n_2 n_3 \log(n_3) + (n_1 + n_2)n_3 r^2 + n_3 r^3)$.

Theoretical guarantees. The following theorem states that ScaledGD algorithm—with proper choices of the tuning parameters, recovers $\mathcal{X}_\star$ with high probability, as long as the fraction of corruptions is not too large and the number of observations is large enough.

Theorem 16 *Let $\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{Y}) = \mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{X}_\star + \mathcal{S}_\star) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where $\mathcal{X}_\star$ is a multi-rank $\boldsymbol{r}$ tensor with μ-incoherence and $\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{S}_\star)$ is αp-sparse. Suppose that the thresholding values $\{\zeta_t\}_{t=0}^\infty$ obey that $\|\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{X}_\star)\|_\infty \leq \zeta_0 \leq 2\|\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{X}_\star)\|_\infty$ and $\zeta_{t+1} = \rho\zeta_t$, $t \geq 1$, for some properly tuned $\zeta_1 = 3\frac{\mu s_r}{n_3\sqrt{n_1 n_2}\ell}\bar{\sigma}_{s_r}(\mathcal{X}_\star)$ and $\frac{1}{5} \leq \eta \leq \frac{1}{2}$, where $\rho = 1 - 0.3\eta$. Then the iterates of ScaledGD*

satisfy

$$\|\mathcal{L}_t *_\Phi \mathcal{R}_t^H - \mathcal{X}_\star\|_F \leq \frac{0.03}{\sqrt{\ell}}\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star),$$

$$\|\mathcal{L}_t *_\Phi \mathcal{R}_t^H - \mathcal{X}_\star\|_\infty \leq 3\rho^t \frac{\mu s_r}{n_3\sqrt{n_1 n_2 \ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star),$$

$$\|\mathcal{S}_t - \mathcal{S}_\star\|_\infty \leq 6\rho^{t-1}\frac{\mu s_r}{n_3\sqrt{n_1 n_2 \ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star) \quad \text{and} \quad \text{supp}(\mathcal{S}_t) \subseteq \text{supp}(\mathcal{P}_\Omega(\mathcal{S}_\star)) \qquad (24)$$

for all $t \geq 1$, as long as the level of corruptions obeys $\alpha \leq \frac{pn_3\sqrt{n_{(2)}n_2}}{10^4 \mu s_r^{1.5}\kappa(n_1+n_2 n_3)} \wedge \frac{pn_3}{10^6 \mu^2 s_r^2 \kappa^2}$ and the probability to observe an entry is high enough, i.e.,

$$p \geq c \log((n_1 \vee n_2)n_3) \max\left(\frac{\mu^{1.5}s_r^{1.5}(n_1 + n_2)}{n_3\sqrt{n_1 n_2}}, \frac{\mu^2 s_r^2 \ell}{n_3\sqrt{(n_1 \wedge n_2)n_3}}, \frac{\mu s_r \kappa^2}{\sqrt{(n_1 \wedge n_2)n_3}}\right)$$

for some sufficiently large constant c.

Theorem 16 establishes that the proposed ScaledGD algorithm (cf. Algorithm 3) finds the ground truth tensor at a constant linear rate, as long as the fraction of corruptions satisfies $\alpha \lesssim \frac{pn_3\sqrt{n_{(2)}n_2}}{\mu s_r^{1.5}\kappa(n_1+n_2 n_3)} \wedge \frac{pn_3}{\mu^2 s_r^2 \kappa^2}$ and the probability of observation satisfies $p \gtrsim \log((n_1 \vee n_2)n_3) \max\left(\frac{\mu^{1.5}s_r^{1.5}(n_1+n_2)}{n_3\sqrt{n_1 n_2}}, \frac{\mu^2 s_r^2 \ell}{n_3\sqrt{(n_1 \wedge n_2)n_3}}, \frac{\mu s_r \kappa^2}{\sqrt{(n_1 \wedge n_2)n_3}}\right)$. Notice that in contrast to Theorem 13, the upper bound of α is proportional to the value of p, which is aligned with the intuition that as we observe more entries in $\mathcal{Y}$, it is still possible to recover $\mathcal{X}_\star$ even for a higher level of corruptions in $\mathcal{X}_\star$.

3.5 Tensor Regression

We now move on to the tensor regression problem. Assume that we have collected a set of observations, given by

$$\boldsymbol{y} = \mathcal{A}(\mathcal{X}_\star) \in \mathbb{R}^m \text{ with } y_i = \langle \mathcal{A}_i, \mathcal{X}_\star \rangle, i = 1, \ldots, m, \qquad (25)$$

where each $\mathcal{A}_i \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ corresponds to the i-th measurement tensor, and $\mathcal{A}(\mathcal{X}) = \{\langle \mathcal{A}_i, \mathcal{X} \rangle\}_{i=1}^m$ is a linear map from $\mathbb{R}^{n_1 \times n_2 \times n_3}$ to $\mathbb{R}^m$. The goal of tensor regression is to recover $\mathcal{X}_\star$ from $\boldsymbol{y}$, especially when the number of observations $m \ll n_1 n_2 n_3$. The preceding discussion helps us pose this problem in terms of the following optimization problem:

$$\min_{\mathcal{F}=[\mathcal{L}^H, \mathcal{R}^H]^H \in \mathbb{R}^{(n_1+n_2) \times r \times n_3}} f(\mathcal{F}) := \frac{1}{2}\|\mathcal{A}(\mathcal{L} *_\Phi \mathcal{R}^H) - \boldsymbol{y}\|_F^2. \qquad (26)$$

The proposed ScaledGD algorithm to minimize (26) is summarized in Algorithm 4. The algorithm starts with an initialization step by applying t-SVD on $\mathcal{A}^*(\boldsymbol{y})$, followed by scaled gradient updates given in (28), where $\mathcal{A}^*(\cdot)$ is the adjoint operator of $\mathcal{A}(\cdot)$, defined as $\mathcal{A}^*(\boldsymbol{y}) = \sum_{i=1}^m y_i \mathcal{A}_i$.

Theoretical guarantees. To understand the performance of ScaledGD for tensor regression, we adopt a standard assumption on the measurement operator $\mathcal{A}(\cdot)$, namely the tensor restricted isometry property (TRIP).

Algorithm 4 ScaledGD for low-rank tensor regression with spectral initialization

---

Input: Linear map $\mathcal{A}(\cdot)$, observation vector $\boldsymbol{y}$, the transformation matrix $\boldsymbol{\Phi}$ associated with the transform L, the tubal rank r, learning rate η, and maximum number of iterations T.

Spectral initialization: Let $\boldsymbol{\mathcal{U}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}_0^H$ be the top-r t-SVD of $\mathcal{A}^*(\boldsymbol{y})$, and set

$$\text{Set} \quad \boldsymbol{\mathcal{L}}_0 = \boldsymbol{\mathcal{U}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0^{\frac{1}{2}} \quad \text{and} \quad \boldsymbol{\mathcal{R}}_0 = \boldsymbol{\mathcal{V}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0^{\frac{1}{2}}. \tag{27}$$

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

$$\boldsymbol{\mathcal{L}}_{t+1} = \boldsymbol{\mathcal{L}}_t - \eta \mathcal{A}^*(\mathcal{A}(\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H) - \boldsymbol{y}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_t^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t)^{-1} \tag{28}$$
$$\boldsymbol{\mathcal{R}}_{t+1} = \boldsymbol{\mathcal{R}}_t - \eta \mathcal{A}^*(\mathcal{A}(\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H) - \boldsymbol{y})^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_t^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_t)^{-1}.$$

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

Definition 17 (TRIP (Rauhut et al., 2017)) *The linear map* $\mathcal{A}(\cdot) : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^m$ *is said to obey the tubal rank-r TRIP with a constant* $\delta_r \in (0, 1)$*, if for all tensors* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *of tubal rank at most r, the following condition holds*

$$(1 - \delta_r)\|\boldsymbol{\mathcal{X}}\|_F^2 \leq \|\mathcal{A}(\boldsymbol{\mathcal{X}})\|_2^2 \leq (1 + \delta_r)\|\boldsymbol{\mathcal{X}}\|_F^2.$$

If $\mathcal{A}(\cdot)$ satisfies tubal rank-$2r$ TRIP with $\delta_{2r} \in (0, 1)$, then for any two tensors $\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of tubal rank at most r, we have

$$(1 - \delta_{2r})\|\boldsymbol{\mathcal{X}}_1 - \boldsymbol{\mathcal{X}}_2\|_F^2 \leq \|\mathcal{A}(\boldsymbol{\mathcal{X}}_1 - \boldsymbol{\mathcal{X}}_2)\|_2^2 \leq (1 + \delta_{2r})\|\boldsymbol{\mathcal{X}}_1 - \boldsymbol{\mathcal{X}}_2\|_F^2.$$

The TRIP under the higher-order singular value decomposition (HOSVD), the tensor train (TT) format, and the Tucker decomposition has been investigated extensively (Rauhut et al., 2017).

Theorem 18 *Suppose that* $\mathcal{A}(\cdot)$ *obeys the $2r$-TRIP with* $\delta_{2r} \leq 0.02/(\sqrt{\frac{s_r}{\ell}}\kappa)$*. If the step size obeys* $0 < \eta \leq \frac{2}{3}$*, then for all $t \geq 0$, the iterates of the ScaledGD method in Algorithm 4 satisfy*

$$\text{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \leq (1 - 0.6\eta)^t 0.1\bar{\sigma}_{s_r}(\boldsymbol{X}_\star) \quad \text{and} \quad \|\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H - \boldsymbol{\mathcal{X}}_\star\|_F \leq (1 - 0.6\eta)^t 0.15\bar{\sigma}_{s_r}(\boldsymbol{X}_\star).$$

Theorem 18 establishes that the distance $\text{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star)$ contracts linearly at a constant rate, as long as $\delta_{2r} \lesssim 1/(\sqrt{\frac{s_r}{\ell}}\kappa)$. To reach ϵ-accuracy, i.e., $\|\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H - \boldsymbol{\mathcal{X}}_\star\|_F \leq \epsilon\bar{\sigma}_{s_r}(\boldsymbol{X}_\star)$, ScaledGD again takes at most $T = \mathcal{O}(\log(1/\epsilon))$ iterations, which is independent of the condition number κ of $\boldsymbol{\mathcal{X}}_\star$.

4. Proof Sketch

In this section, we provide some intuitions and sketch the proof of our main theorems.

4.1 ScaledGD for Tensor Factorization

To shed light on why ScaledGD is robust to ill-conditioning, we first consider the problem of factorizing a tensor $\boldsymbol{\mathcal{X}}_\star$ into two low-rank factors:

$$\min_{\boldsymbol{\mathcal{F}}=[\boldsymbol{\mathcal{L}}^H, \boldsymbol{\mathcal{R}}^H]^H \in \mathbb{R}^{(n_1+n_2) \times r \times n_3}} f(\boldsymbol{\mathcal{F}}) := \frac{1}{2} \|\boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}^H - \boldsymbol{\mathcal{X}}_\star\|_F^2. \tag{29}$$

Recalling the update rule (4), ScaledGD proceeds as follows:

$$\boldsymbol{\mathcal{L}}_{t+1} = \boldsymbol{\mathcal{L}}_t - \eta(\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H - \boldsymbol{\mathcal{X}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_t^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t)^{-1},$$
$$\boldsymbol{\mathcal{R}}_{t+1} = \boldsymbol{\mathcal{R}}_t - \eta(\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H - \boldsymbol{\mathcal{X}}_\star)^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_t^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_t)^{-1}, \tag{30}$$

Due to property (8), problem (29) can be transformed into the transformed domain by solving the following problem:

$$\min_{\widetilde{\boldsymbol{F}}=[\bar{\boldsymbol{L}}^H, \bar{\boldsymbol{R}}^H]^H \in \mathbb{C}^{(n_1+n_2)n_3 \times n_3 r}} f(\bar{\boldsymbol{L}}, \bar{\boldsymbol{R}}) := \frac{1}{2\ell} \|\bar{\boldsymbol{L}} \bar{\boldsymbol{R}}^H - \bar{\boldsymbol{X}}_\star\|_F^2, \tag{31}$$

where $\bar{\boldsymbol{L}} = \texttt{bdiag}(\bar{\boldsymbol{\mathcal{L}}})$ as defined in (7). Note that we use $\widetilde{\boldsymbol{F}} = \begin{bmatrix} \bar{\boldsymbol{L}} \\ \bar{\boldsymbol{R}} \end{bmatrix}$ here to denote the vertical concatenation of $\bar{\boldsymbol{L}}$ and $\bar{\boldsymbol{R}}$, which differs from the matrix $\bar{\boldsymbol{F}} = \texttt{bdiag}(\bar{\boldsymbol{\mathcal{F}}})$ only by a permutation of rows, where $\bar{\boldsymbol{\mathcal{F}}} = L(\boldsymbol{\mathcal{F}})$ with $\boldsymbol{\mathcal{F}} = \begin{bmatrix} \boldsymbol{\mathcal{L}} \\ \boldsymbol{\mathcal{R}} \end{bmatrix}$. This is now a standard matrix least-squares problem, up to a scalar. The gradients of $f(\widetilde{\boldsymbol{F}})$ in (31) with respect to $\bar{\boldsymbol{L}}$ and $\bar{\boldsymbol{R}}$ are given by

$$\nabla_{\bar{\boldsymbol{L}}} f(\widetilde{\boldsymbol{F}}) = \frac{1}{\ell} (\bar{\boldsymbol{L}} \bar{\boldsymbol{R}}^H - \bar{\boldsymbol{X}}_\star) \bar{\boldsymbol{R}} \quad \text{and} \quad \nabla_{\bar{\boldsymbol{R}}} f(\widetilde{\boldsymbol{F}}) = \frac{1}{\ell} (\bar{\boldsymbol{L}} \bar{\boldsymbol{R}}^H - \bar{\boldsymbol{X}}_\star)^H \bar{\boldsymbol{L}},$$

which allows for the computation of the Hessian with respect to $\bar{\boldsymbol{L}}$ and $\bar{\boldsymbol{R}}$. When written in terms of the vectorized variables, the Hessians are expressed as

$$\nabla_{\bar{\boldsymbol{L}}, \bar{\boldsymbol{L}}}^2 f(\widetilde{\boldsymbol{F}}) = \frac{1}{\ell} (\bar{\boldsymbol{R}}^H \bar{\boldsymbol{R}}) \otimes \boldsymbol{I}_{n_1 n_3} \quad \text{and} \quad \nabla_{\bar{\boldsymbol{R}}, \bar{\boldsymbol{R}}}^2 f(\widetilde{\boldsymbol{F}}) = \frac{1}{\ell} (\bar{\boldsymbol{L}}^H \bar{\boldsymbol{L}}) \otimes \boldsymbol{I}_{n_2 n_3},$$

where $\otimes$ denotes the Kronecker product. Let $\texttt{vec}(\boldsymbol{A})$ denote the vectorization of a matrix $\boldsymbol{A}$. Viewed in the vectorized form, it is not difficult to check that the ScaledGD update (30) is equivalent to approximating the Hessian of the loss function (31) by only keeping its diagonal blocks, i.e.,

$$\texttt{vec}(\widetilde{\boldsymbol{F}}_{t+1}) = \texttt{vec}(\widetilde{\boldsymbol{F}}_t) - \frac{\eta}{\ell} \begin{bmatrix} \nabla_{\bar{\boldsymbol{L}}, \bar{\boldsymbol{L}}}^2 f(\widetilde{\boldsymbol{F}}_t) & \boldsymbol{0} \\ \boldsymbol{0} & \nabla_{\bar{\boldsymbol{R}}, \bar{\boldsymbol{R}}}^2 f(\widetilde{\boldsymbol{F}}_t) \end{bmatrix}^{-1} \texttt{vec}(\nabla_{\widetilde{\boldsymbol{F}}} f(\widetilde{\boldsymbol{F}}_t)).$$

Under this formulation, the ScaledGD update corresponds to a Newton-type update up to a constant step size, yielding a natural quasi-Newton interpretation where the preconditioner is designed as the inverse of the diagonal approximation of the Hessian. Compared with vanilla gradient descent, ScaledGD exploits additional curvature information through a structured scaling of the gradient. This quasi-Newton interpretation provides a natural explanation for its improved convergence behavior, especially in ill-conditioned regimes.

17

Theoretical guarantees for tensor factorization. The following theorem, whose proof can be found in Appendix B, formally establishes that as long as initialization is not too far from the ground truth, $\text{dist}(\mathcal{F}_t, \mathcal{F}_\star)$ will contract at a constant linear rate for the tensor factorization problem.

Theorem 19 *Suppose that the initialization $\mathcal{F}_0$ satisfies $\text{dist}(\mathcal{F}_0, \mathcal{F}_\star) \leq \frac{0.1}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star)$. If the step size obeys $0 < \eta \leq \frac{2}{3}$, then for all $t \geq 0$, the iterates of the ScaledGD method in (30) satisfy*

$$\text{dist}(\mathcal{F}_t, \mathcal{F}_\star) \leq (1 - 0.7\eta)^t \frac{0.1}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star) \quad \text{and} \quad \|\mathcal{L}_t *_\Phi \mathcal{R}_t^H - \mathcal{X}_\star\|_F \leq (1 - 0.7\eta)^t \frac{0.15}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star).$$

4.2 Proof Outline for Tensor RPCA

The proof of Theorem 13 is inductive in nature, where we aim to establish the following induction hypothesis at all the iterations:

$$\text{dist}(\mathcal{F}_t, \mathcal{F}_\star) \leq \frac{\epsilon}{\sqrt{\ell}}\rho^t\bar{\sigma}_{s_r}(\mathcal{X}_\star), \quad \text{and}$$

$$\sqrt{n_1}\|(\mathcal{L}_t *_\Phi \mathcal{Q}_t - \mathcal{L}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|(\mathcal{R}_t *_\Phi \mathcal{Q}_t^{-H} - \mathcal{R}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{n_3\ell}}\rho^t\bar{\sigma}_{s_r}(\mathcal{X}_\star).$$

The following two lemmas, establishes the induction hypothesis for both the induction case and the base case. We start by outlining the local contraction of the proposed Algorithm 1.

Lemma 20 (Local contraction) *Let $\mathcal{Y} = \mathcal{X}_\star + \mathcal{S}_\star \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where $\mathcal{X}_\star = \mathcal{L}_\star *_\Phi \mathcal{R}_\star^H$ is a multi-rank r tensor with μ-incoherence and $\mathcal{S}_\star$ is α-sparse. Let $\mathcal{Q}_t$ be the optimal alignment tensor between $\begin{bmatrix} \mathcal{L}_t \\ \mathcal{R}_t \end{bmatrix}$ and $\begin{bmatrix} \mathcal{L}_\star \\ \mathcal{R}_\star \end{bmatrix}$. Under the assumption that $\alpha \leq \frac{n_3\sqrt{n_{(2)}n_2}}{10^4\mu s_r^{1.5}(n_1+n_2n_3)}$, if the spectral initialization obeys the conditions*

$$\text{dist}(\mathcal{F}_0, \mathcal{F}_\star) \leq \frac{\epsilon}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star),$$

$$\sqrt{n_1}\|(\mathcal{L}_0 *_\Phi \mathcal{Q}_0 - \mathcal{L}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|(\mathcal{R}_0 *_\Phi \mathcal{Q}_0^{-H} - \mathcal{R}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{n_3\ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star)$$

with $\epsilon = 0.02$, then by setting the thresholding values ζ_t in Theorem 13 and the step size $\frac{1}{4} \leq \eta \leq \frac{8}{9}$, the iterates of Algorithm 1 satisfy

$$\text{dist}(\mathcal{F}_t, \mathcal{F}_\star) \leq \frac{\epsilon}{\sqrt{\ell}}\rho^t\bar{\sigma}_{s_r}(\mathcal{X}_\star),$$

$$\sqrt{n_1}\|(\mathcal{L}_t *_\Phi \mathcal{Q}_t - \mathcal{L}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|(\mathcal{R}_t *_\Phi \mathcal{Q}_t^{-H} - \mathcal{R}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{n_3\ell}}\rho^t\bar{\sigma}_{s_r}(\mathcal{X}_\star),$$

where the convergence rate $\rho = 1 - 0.6\eta$.

The following lemma ensures that the spectral initialization satisfies the distance and incoherence conditions.

Lemma 21 (Guaranteed initialization) *Let $\mathcal{Y} = \mathcal{X}_\star + \mathcal{S}_\star \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where $\mathcal{X}_\star = \mathcal{L}_\star *_{\Phi} \mathcal{R}_\star^H$ is a multi-rank r tensor with μ-incoherence and $\mathcal{S}_\star$ is α-sparse. Under the assumption that $\alpha \le \frac{c_0}{\mu s_r^{1.5} \kappa \frac{n_1+n_2 n_3}{n_3 \sqrt{n_1 n_2}}} \wedge \frac{c_0^2 n_3}{\mu^2 s_r^2 \kappa^2}$ for some small positive constant $c_0 \le 0.06$ and the choice of the thresholding value $\|\mathcal{X}_\star\|_\infty \le \zeta_0 \le 2\|\mathcal{X}_\star\|_\infty$, the spectral initialization satisfies*

$$\operatorname{dist}(\mathcal{F}_0, \mathcal{F}_\star) \le \frac{5c_0}{\sqrt{\ell}} \bar{\sigma}_{s_r}(\mathcal{X}_\star) \quad \text{and}$$

$$\sqrt{n_1}\|(\mathcal{L}_0 *_{\Phi} \mathcal{Q}_0 - \mathcal{L}_\star) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|(\mathcal{R}_0 *_{\Phi} \mathcal{Q}_0^{-H} - \mathcal{R}_\star) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \le \sqrt{\frac{\mu s_r}{n_3 \ell}} \bar{\sigma}_{s_r}(\mathcal{X}_\star),$$

where $\mathcal{Q}_0$ be the optimal alignment tensor between $\begin{bmatrix} \mathcal{L}_0 \\ \mathcal{R}_0 \end{bmatrix}$ and $\begin{bmatrix} \mathcal{L}_\star \\ \mathcal{R}_\star \end{bmatrix}$.

The proofs of the above two lemmas are provided in Appendix C. We also present the following lemma that verifies the selection of thresholding value is indeed effective.

Lemma 22 (Cai et al. (2021), Lemma 5) *At the $(t+1)$-th iteration of Algorithm 1, taking the thresholding value $\zeta_{t+1} = \|\mathcal{X}_\star - \mathcal{X}_t\|_\infty$ gives*

$$\|\mathcal{S}_\star - \mathcal{S}_{t+1}\|_\infty \le 2\|\mathcal{X}_\star - \mathcal{X}_t\|_\infty \quad \text{and} \quad \operatorname{supp}(\mathcal{S}_{t+1}) \subseteq \operatorname{supp}(\mathcal{S}_\star).$$

4.3 Proof Outline for Tensor Completion

We start with the following lemma that ensures the scaled projection in (19) satisfies both non-expansiveness and incoherence under the scaled metric.

Lemma 23 *Suppose that $\mathcal{X}_\star$ is μ-incoherent, and $\operatorname{dist}(\widetilde{\mathcal{F}}, \mathcal{F}_\star) \le \frac{\epsilon}{\sqrt{\ell}} \bar{\sigma}_{s_r}(\mathcal{X}_\star)$ for some $\epsilon < 1$. Set $\varsigma \ge (1+\epsilon)\sqrt{\frac{\mu s_r}{n_3 \ell}} \bar{\sigma}_1(\mathcal{X}_\star)$, then $\mathcal{P}_\varsigma(\widetilde{\mathcal{F}})$ satisfies the non-expansiveness*

$$\operatorname{dist}(\mathcal{P}_\varsigma(\widetilde{\mathcal{F}}), \mathcal{F}_\star) \le \operatorname{dist}(\widetilde{\mathcal{F}}, \mathcal{F}_\star),$$

and the incoherence condition

$$\sqrt{n_1}\|\mathcal{L} *_{\Phi} \mathcal{R}^H\|_{2,\infty} \vee \sqrt{n_2}\|\mathcal{R} *_{\Phi} \mathcal{L}^H\|_{2,\infty} \le \varsigma.$$

Our next lemma guarantees the fast local convergence of Algorithm 2 as long as the sample complexity is large enough and the parameter ς is set properly.

Lemma 24 *Suppose that $\mathcal{X}_\star$ is μ-incoherent, and $p \ge c\left(\frac{\mu s_r(n_1+n_2)\log((n_1 \vee n_2)n_3)}{n_1 n_2 n_3} \vee \frac{\mu^2 s_r^2 \kappa^4 \ell \log((n_1 \vee n_2)n_3)}{(n_1 \wedge n_2)n_3^2}\right)$ for some sufficiently large constant c. Set the projection radius as $\varsigma \ge c_\varsigma \sqrt{\frac{\mu s_r}{n_3 \ell}} \bar{\sigma}_1(\mathcal{X}_\star)$ for some constant $c_\varsigma \ge 1.02$. Under an event G which happens with high probability, if the t-th iterate satisfies $\operatorname{dist}(\mathcal{F}_t, \mathcal{F}_\star) \le \frac{0.02}{\sqrt{\ell}} \bar{\sigma}_{s_r}(\mathcal{X}_\star)$, and the incoherence condition*

$$\sqrt{n_1}\|\mathcal{L}_t *_{\Phi} \mathcal{R}_t^H\|_{2,\infty} \vee \sqrt{n_2}\|\mathcal{R}_t *_{\Phi} \mathcal{L}_t^H\|_{2,\infty} \le \varsigma,$$

19

then $\|\mathcal{L}_t *_{\boldsymbol{\Phi}} \mathcal{R}_t^H - \mathcal{X}_\star\|_F \leq 1.5\mathrm{dist}(\mathcal{F}_t, \mathcal{F}_\star)$. In addition, if the step size obeys $0 < \eta \leq \frac{2}{3}$, then the $(t+1)$-th iterate $\mathcal{F}_{t+1}$ of the ScaledGD method in (21) of Algorithm 2 satisfies

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

and the incoherence condition

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

Therefore, if we can find an initialization that is close to the ground truth and satisfies the incoherence condition, Lemma 24 guarantees that the iterates of ScaledGD remain incoherent and converge linearly.

Lemma 25 *Suppose that $\mathcal{X}_\star$ is μ-incoherent, then with high probability, the spectral initialization before projection* $\widetilde{\mathcal{F}}_0 = \begin{bmatrix} \mathcal{U}_0 *_{\boldsymbol{\Phi}} \mathcal{G}_0^{\frac{1}{2}} \\ \mathcal{V}_0 *_{\boldsymbol{\Phi}} \mathcal{G}_0^{\frac{1}{2}} \end{bmatrix}$ *in (20) satisfies*

$$\mathrm{dist}(\widetilde{\mathcal{F}}_0, \mathcal{F}_\star) \leq c\left(\frac{\mu s_r \log((n_1 \vee n_2)n_3)}{pn_3\sqrt{n_1 n_2}} + \sqrt{\frac{\mu s_r \log((n_1 \vee n_2)n_3)}{p(n_1 \wedge n_2)n_3}}\right)5\sqrt{\frac{s_r}{\ell}}\kappa\bar{\sigma}_{s_r}(\mathcal{X}_\star).$$

Therefore, as long as $p \geq c\mu s_r^2 \kappa^2 \log((n_1 \vee n_2)n_3)/((n_1 \wedge n_2)n_3)$ for some sufficiently large constant c, the initial distance satisfies $\mathrm{dist}(\widetilde{\mathcal{F}}_0, \mathcal{F}_\star) \leq \frac{0.02}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star)$. We can then invoke Lemma 23 to have that $\mathcal{F}_0 = \mathcal{P}_\varsigma(\widetilde{\mathcal{F}}_0)$ meets the conditions required in Lemma 24, which further enables us to repetitively apply Lemma 23 to finish the proof of Theorem 15. The proofs of the three supporting lemmas can be found in Appendix D.

4.4 Proof Outline for Robust Tensor Completion

We first present the following two lemmas of local linear convergence and guaranteed initialization, whose proofs are deferred to Appendix E. Then the claims in Theorem 16 follow immediately in the same fashion as the proof of Theorem 13 by setting $c_0 \leq 0.001$.

Lemma 26 (Local contraction) *Let $\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{Y}) = \mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{X}_\star + \mathcal{S}_\star) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where $\mathcal{X}_\star = \mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H$ is a multi-rank $\boldsymbol{r}$ tensor with μ-incoherence and $\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{S}_\star)$ is αp-sparse. Let $\mathcal{Q}_t$ be the optimal alignment tensor between $\begin{bmatrix} \mathcal{L}_t \\ \mathcal{R}_t \end{bmatrix}$ and $\begin{bmatrix} \mathcal{L}_\star \\ \mathcal{R}_\star \end{bmatrix}$. Under the assumption that $\alpha \leq \frac{pn_3\sqrt{n_{(2)}n_2}}{10^4 \mu s_r^{1.5}(n_1 + n_2 n_3)}$, where $p \geq c\left(\frac{\mu^{1.5}s_r^{1.5}(n_1 + n_2)\log((n_1 \vee n_2)n_3)}{n_3\sqrt{n_1 n_2}} \vee \frac{\mu^2 s_r^2 \ell \log((n_1 \vee n_2)n_3)}{n_3\sqrt{(n_1 \wedge n_2)n_3}}\right)$ for some sufficiently large constant c. If the spectral initialization obeys the conditions*

$$\mathrm{dist}(\mathcal{F}_0, \mathcal{F}_\star) \leq \frac{\epsilon}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star),$$

$$\sqrt{n_1}\|(\mathcal{L}_0 *_{\boldsymbol{\Phi}} \mathcal{Q}_0 - \mathcal{L}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|(\mathcal{R}_0 *_{\boldsymbol{\Phi}} \mathcal{Q}_0^{-H} - \mathcal{R}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{n_3 \ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star)$$

with $\epsilon = 0.02$, then by setting the thresholding values ζ_t in Theorem 16 and the step size $\frac{1}{5} \leq \eta \leq \frac{1}{2}$, the iterates of Algorithm 3 satisfy

$$\text{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \leq \frac{\epsilon}{\sqrt{\ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star),$$

$$\sqrt{n_1} \|(\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2} \|(\boldsymbol{\mathcal{R}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star),$$

where the convergence rate $\rho = 1 - 0.3\eta$.

Lemma 27 (Guaranteed initialization) *Let $\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{Y}}) = \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star + \boldsymbol{\mathcal{S}}_\star) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where $\boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\star^H$ is a multi-rank $\boldsymbol{r}$ tensor with μ-incoherence and $\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{S}}_\star)$ is αp-sparse. Under the assumption that $\alpha \leq \frac{c_0 p}{\mu s_r^{1.5} \kappa \frac{n_1 + n_2 n_3}{n_3 \sqrt{n_1 n_2}}} \wedge \frac{c_0^2 p n_3}{\mu^2 s_r^2 \kappa^2}$ for some small positive constant $c_0 \leq 0.05$, where $p \geq c\mu s_r \kappa^2 \log((n_1 \vee n_2)n_3)/\sqrt{(n_1 \wedge n_2)n_3}$ for some sufficiently large constant c. Given the choice of the thresholding value $\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star)\|_\infty \leq \zeta_0 \leq 2\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star)\|_\infty$, the spectral initialization satisfies*

$$\text{dist}(\boldsymbol{\mathcal{F}}_0, \boldsymbol{\mathcal{F}}_\star) \leq \frac{5c_0 + 0.01}{\sqrt{\ell}} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) \quad \text{and}$$

$$\sqrt{n_1} \|(\boldsymbol{\mathcal{L}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_0 - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2} \|(\boldsymbol{\mathcal{R}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_0^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{n_3 \ell}} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star),$$

where $\boldsymbol{\mathcal{Q}}_0$ be the optimal alignment tensor between $\begin{bmatrix} \boldsymbol{\mathcal{L}}_0 \\ \boldsymbol{\mathcal{R}}_0 \end{bmatrix}$ and $\begin{bmatrix} \boldsymbol{\mathcal{L}}_\star \\ \boldsymbol{\mathcal{R}}_\star \end{bmatrix}$.

4.5 Proof Outline for Tensor Regression

Now we turn to the proof outline for tensor regression (cf. Theorem 18). Leveraging the TRIP of $\mathcal{A}(\cdot)$, we can establish the following local convergence guarantee of ScaledGD, as long as the iterates are close to the ground truth.

Lemma 28 *Suppose that $\mathcal{A}(\cdot)$ obeys the $2r$-TRIP with $\delta_{2r} \leq 0.02$. If the t-th iterate satisfies $\text{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \leq 0.1\bar{\sigma}_{s_r}(\boldsymbol{X}_\star)$, then $\|\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H - \boldsymbol{\mathcal{X}}_\star\|_F \leq 1.5\text{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star)$. In addition, if the step size obeys $0 < \eta \leq \frac{2}{3}$, then the $(t+1)$-th iterate $\boldsymbol{\mathcal{F}}_{t+1}$ of the ScaledGD method in (28) of Algorithm 4 satisfies*

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

To establish the induction hypothesis, we still need to check the quality of the spectral initialization, for which we have the following lemma.

Lemma 29 *Suppose that $\mathcal{A}(\cdot)$ obeys the $2r$-TRIP with a constant δ_{2r}. Then the spectral initialization in (27) for low-rank tensor regression satisfies*

$$\text{dist}(\boldsymbol{\mathcal{F}}_0, \boldsymbol{\mathcal{F}}_\star) \leq 5\delta_{2r} \sqrt{\frac{s_r}{\ell}} \kappa \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).$$

As a result, setting $\delta_{2r} \leq 0.02/(\sqrt{\frac{s_r}{\ell}}\kappa)$ as specified in Theorem 18, the initial distance satisfies $\text{dist}(\boldsymbol{\mathcal{F}}_0, \boldsymbol{\mathcal{F}}_\star) \leq 0.1\bar{\sigma}_{s_r}(\boldsymbol{X}_\star)$, allowing us to invoke Lemma 28 recursively. The proof of Theorem 18 is then complete. The proofs of Lemma 28 and Lemma 29 can be found in Appendix F.

5. Numerical Experiments

In this section, we present several experimental results demonstrating the effectiveness of our proposed methods. We first provide numerical experiments to corroborate our theoretical findings. Then we evaluate the performance of our algorithms by focusing on video denoising and background initialization tasks. All experiments are performed in Matlab with an AMD Ryzen 9 5950X 3.40GHz CPU and 64GB RAM.

5.1 Synthetic Data Experiments

We compare the iteration complexity of ScaledGD with vanilla gradient descent (GD). For fair comparison, both algorithms start from the same spectral initialization, and the update rule of GD is given by

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

where $\eta_{\mathrm{GD}} = \eta/\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star)$ stands for the step size for vanilla GD.

In this experiment, we adopt two invertible linear transforms: (a) Discrete Fourier Transform (DFT) and (b) Discrete Cosine Transform (DCT). For simplicity, we set $n_1 = n_2 = n$. We generate the ground truth tensor $\boldsymbol{\mathcal{X}}_\star \in \mathbb{R}^{n \times n \times n_3}$ as follows. We first generate an $n \times r \times n_3$ tensor with i.i.d. random signs, and take its r left singular tensors as $\boldsymbol{\mathcal{U}}_\star$, and similarly for $\boldsymbol{\mathcal{V}}_\star$. The diagonal entries in each frontal slice of the f-diagonal tensor $\bar{\boldsymbol{\mathcal{G}}}_\star$ are set to be linearly distributed from 1 to $1/\kappa$. Then the underlying low-rank tensor is generated by $\boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{U}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}_\star^H$, which has the specified condition number κ and tubal rank r. We consider the following four low-rank tensor estimation tasks:

- *Tensor RPCA.* We generate the sparse corruption tensor by uniformly and independently sampling α-fraction of the entries as the non-zero locations of $\boldsymbol{\mathcal{S}}_\star$ with $\alpha = 0.1$. The magnitudes of the non-zero entries of $\boldsymbol{\mathcal{S}}_\star$ are sampled i.i.d. from the uniform distribution over the interval of $[-\mathbb{E}[|\boldsymbol{\mathcal{X}}_\star|_{i,j,k}], \mathbb{E}[|\boldsymbol{\mathcal{X}}_\star|_{i,j,k}]]$. The observation tensor is $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}_\star + \boldsymbol{\mathcal{S}}_\star + \boldsymbol{\mathcal{W}}$, where $\boldsymbol{\mathcal{W}}_{i,j,k} \sim \mathcal{N}(0, \sigma_w^2)$ composed of i.i.d. Gaussian entries.

- *Tensor completion.* We assume random Bernoulli observations, where each entry of $\boldsymbol{\mathcal{X}}_\star$ is observed with probability $p = 0.4$ independently. The observation is $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star + \boldsymbol{\mathcal{W}})$, where $\boldsymbol{\mathcal{W}}_{i,j,k} \sim \mathcal{N}(0, \sigma_w^2)$ is composed of i.i.d. Gaussian entries. Moreover, we perform the scaled gradient updates without projections.

- *Robust tensor completion.* The problem formulation is stated in Section 3.4. We first generate the corruption using a sparse tensor $\boldsymbol{\mathcal{S}}_\star \in \mathscr{S}_\alpha$ with $\alpha = 0.1$. The observation is generated by $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{X}}_\star + \boldsymbol{\mathcal{S}}_\star)$, where $\boldsymbol{\Omega}$ is generated according to the Bernoulli model with probability $p = 0.6$.

- *Tensor regression.* The problem formulation is detailed in Section 3.5. Here, we collect $m = 5nn_3r$ measurements in the form of $y_i = \langle \boldsymbol{\mathcal{A}}_i, \boldsymbol{\mathcal{X}}_\star \rangle$, in which the measurement tensors $\boldsymbol{\mathcal{A}}_i$ are generated with i.i.d. Gaussian entries with zero mean and variance $1/m$.

Figure 1: The relative errors of ScaledGD and vanilla GD with respect to the iteration count under different condition numbers $\kappa = 1, 5, 10, 20$ for (a/b) tensor RPCA, (c/d) tensor completion, (e/f) robust tensor completion, and (g/h) tensor regression.

We simply set $\eta = 0.4$ for the robust tensor completion problem and fix $\eta = 0.5$ for both ScaledGD and vanilla GD for the other three problems (see Figure 3 for justifications). The hyperparameters for tensor RPCA and robust tensor completion are set as follows: $\zeta_0 = 0.5$, $\zeta_1 = 0.5$, and $\rho = 0.95$. For simplicity, we set $n_3 = n = 100$ and $r = 10$ for the first three problems and set $n = 50$, $n_3 = 20$ and $r = 5$ for the tensor regression problem.

We first illustrate the convergence performance under noise-free observations, i.e., $\mathcal{W} = \mathbf{0}$. Figure 1 shows the relative reconstruction error $\|\boldsymbol{\mathcal{X}}_t - \boldsymbol{\mathcal{X}}_\star\|_F / \|\boldsymbol{\mathcal{X}}_\star\|_F$ with respect to the iteration count t for the four problems under different condition numbers $\kappa = 1, 5, 10, 20$ using the two transforms. For the tensor RPCA and robust tensor completion problems, ScaledGD converges at the same rate as vanilla GD under good conditioning (e.g., $\kappa = 1, 5$ for tensor RPCA and $\kappa = 1$ for robust tensor completion); under ill-conditioning, i.e., when

(a) Tensor RPCA, DFT
$n = 100$, $\alpha = 0.1$

(b) Tensor RPCA, DCT
$n = 100$, $\alpha = 0.1$

(c) Tensor Completion, DFT
$n = 100$, $p = 0.4$

(d) Tensor Completion, DCT
$n = 100$, $p = 0.4$

(e) RTC, DFT
$n = 100$, $\alpha = 0.1$, $p = 0.6$

(f) RTC, DCT
$n = 100$, $\alpha = 0.1$, $p = 0.6$

(g) Tensor Regression, DFT
$n = 50$, $m = 25000$

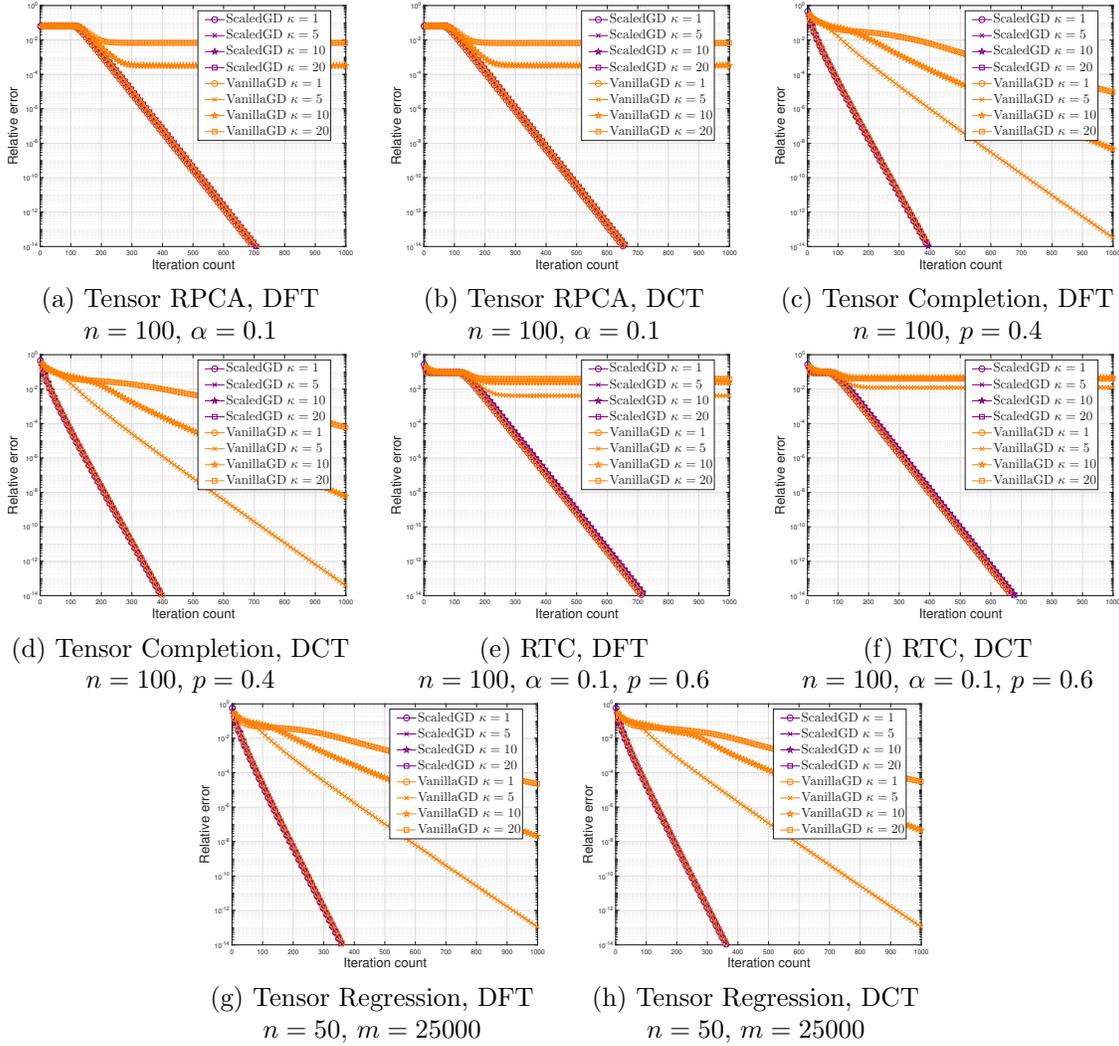(h) Tensor Regression, DCT
$n = 50$, $m = 25000$
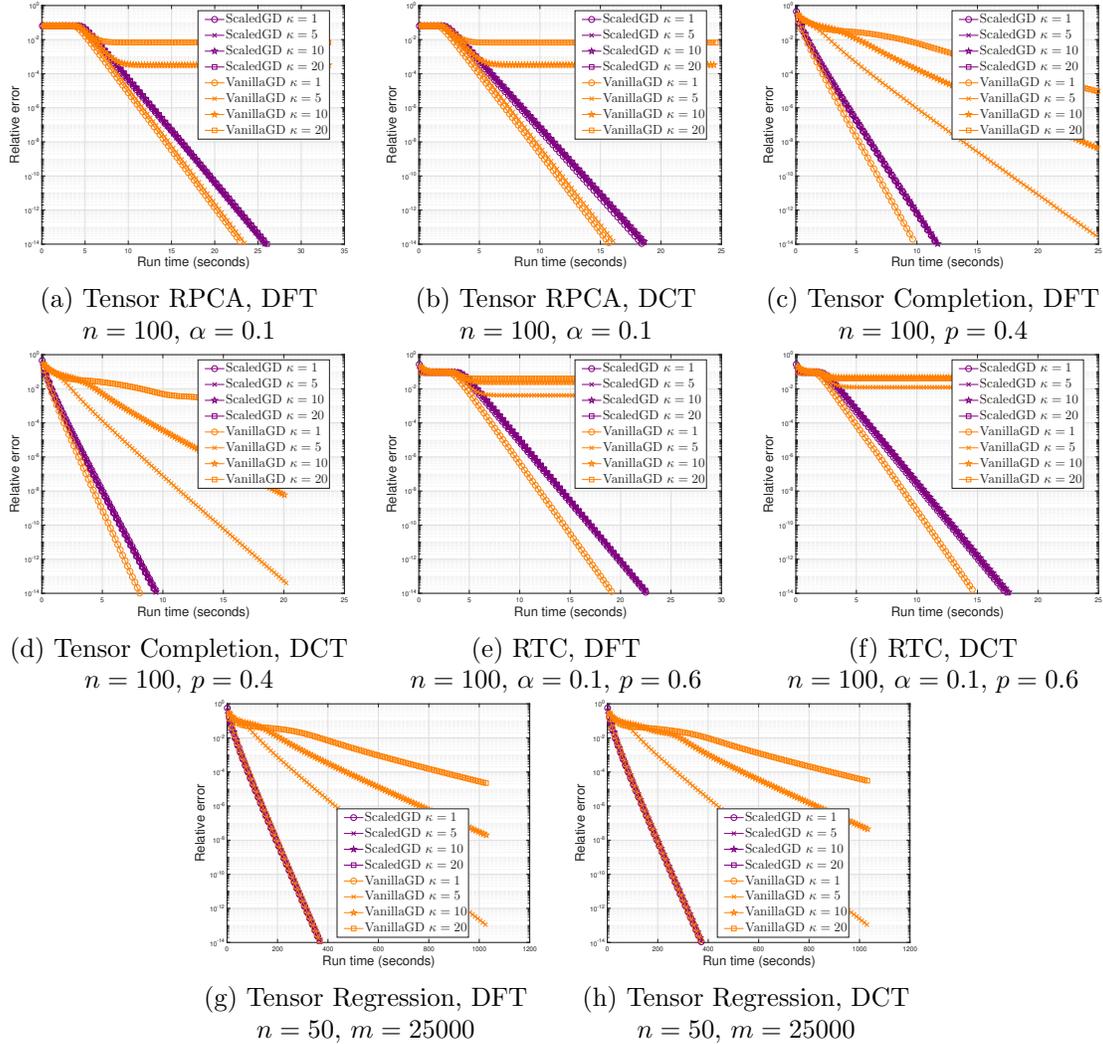
Figure 2: The relative errors of ScaledGD and vanilla GD with respect to run time (in seconds) under different condition numbers $\kappa = 1, 5, 10, 20$ for (a/b) tensor RPCA, (c/d) tensor completion, (e/f) robust tensor completion, and (g/h) tensor regression.

κ is large, ScaledGD converges linearly with a rate that is independent of κ, while vanilla GD does not converge because the relative error stays above 10^{-4}. For the tensor completion and tensor regression problems, we can see that ScaledGD converges rapidly at a rate independent of the condition number, and matches the fastest rate of vanilla GD with perfect conditioning $\kappa = 1$. We plot the relative reconstruction errors of ScaledGD and vanilla GD for the four problems with respect to the algorithm running time (in seconds) under different condition numbers $\kappa = 1, 5, 10, 20$ in Figure 2. We again observe the advantage of ScaledGD over vanilla GD for the tensor RPCA and robust tensor completion problems under ill-conditioning because vanilla GD does not converge in this scenario. For the tensor completion problem, although vanilla GD runs slightly faster than ScaledGD when $\kappa = 1$, the convergence rate of vanilla GD deteriorates quickly with the increase of κ. As such,

(a) Tensor RPCA, DFT

(b) Tensor RPCA, DCT

(c) Tensor Completion, DFT

(d) Tensor Completion, DCT
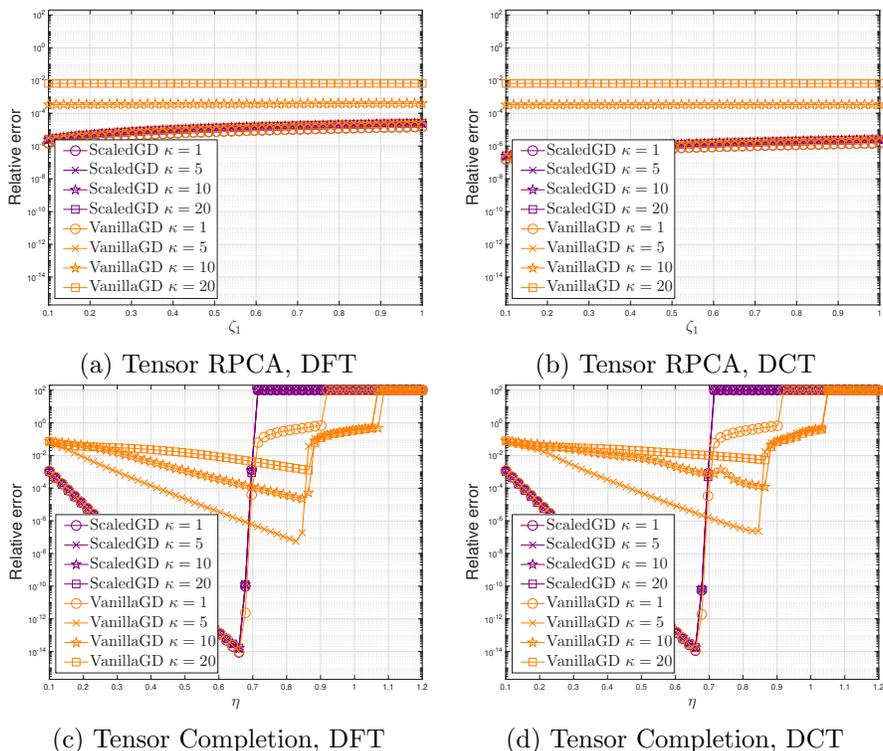
Figure 3: (a) and (b) show the relative errors of ScaledGD and vanilla GD after 300 iterations with respect to different ζ_1's from 0.1 to 1 under different condition numbers $\kappa = 1, 5, 10, 20$ for tensor RPCA with $n = 100$, $r = 10$, and $\alpha = 0.1$. (c) and (d) show the relative errors of ScaledGD and vanilla GD after 300 iterations with respect to different step sizes η from 0.1 to 1.2 under different condition numbers $\kappa = 1, 5, 10, 20$ for tensor completion with $n = 100$, $r = 10$, and $p = 0.4$.

under ill-conditioning, the computational burdens can be substantially increased for vanilla GD compared to ScaledGD. Finally, Figure 2g and Figure 2h suggest that ScaledGD can be as fast as vanilla GD even under perfect conditioning for tensor regression, while vanilla GD turns out to be much slower than ScaledGD when κ increases.

Next, we study the sensitivity of the convergence behavior of ScaledGD with respect to the choice of the hyperparameters. We first examine the effect of ζ_1 on the convergence speeds of ScaledGD and vanilla GD for tensor RPCA by fixing $\zeta_0 = 0.5$, $\eta = 0.5$, and $\rho = 0.95$. We run both algorithms for at most 300 iterations (the algorithm is terminated if the relative error exceeds 10^2). As indicated in Figure 3a and Figure 3b, ScaledGD achieves exact recovery over a wide range of ζ_1 values, i.e., the relative error is always below 10^{-5} and the recovery performance remains unaffected by the condition number. In contrast, vanilla GD cannot achieve exact recovery when κ gets large. Figure 3c and Figure 3d illustrate the convergence speeds of ScaledGD and vanilla GD under different step sizes for tensor completion, where we again run both algorithms for at most 300 iterations. It can be seen that ScaledGD outperforms vanilla GD over a large range of step sizes, even when the step size of vanilla GD is optimized for its performance. Hence, our choice of $\eta = 0.5$ in the

(a) Tensor RPCA, DFT   (b) Tensor RPCA, DCT

(c) Tensor completion, DFT   (d) Tensor completion, DCT
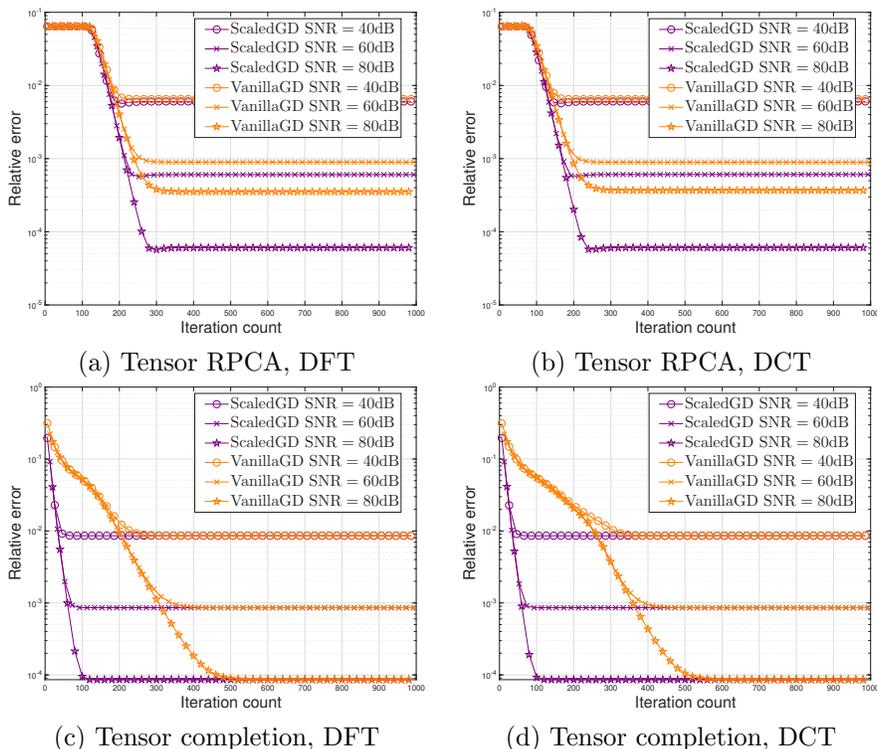
Figure 4: The relative errors of ScaledGD and vanilla GD with respect to the iteration count under the condition number $\kappa = 10$ and signal-to-noise ratios SNR $= 40, 60, 80$dB for (a/b) tensor RPCA and (c/d) tensor completion.

previous tensor completion experiments for the comparison between ScaledGD and vanilla GD is reasonable.

Finally, we examine the performance of ScaledGD for tensor RPCA and tensor completion under Gaussian noisy observations. We denote the signal-to-noise ratio as SNR $:= 10 \log_{10} \frac{\|\boldsymbol{\mathcal{X}}_\star\|_F^2}{n^2 n_3 \sigma_w^2}$ 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 4 under the condition number $\kappa = 10$ and various SNR $= 40, 60, 80$dB. For tensor RPCA, ScaledGD achieves smaller error compared to vanilla GD. For tensor completion, both methods achieve the same statistical error eventually, but ScaledGD converges much faster. In addition, the convergence speeds of ScaledGD are irrespective of the noise levels.

5.2 Real-World Applications

In this section, we apply ScaledGD for tensor RPCA and compare the performance of ScaledGD with other tensor RPCA algorithms, including TRPCA (Lu et al., 2020; Lu, 2021), ScaledGD using Tucker decomposition (ScaledGD-Tucker) (Dong et al., 2023b) and EAPT (Qiu et al., 2022) on real datasets. We again use two different linear transforms for ScaledGD, TRPCA and EAPT in all experiments, i.e., DFT and DCT. The corresponding methods of ScaledGD are called ScaledGD-DFT and ScaledGD-DCT for short, and this

Table 1: Comparison of video data denoising performance with different levels of corruptions. The best result is shown in bold.

Methods	$\alpha = 0.05$			$\alpha = 0.1$		
	Time(sec)	RSE	PSNR	Time(sec)	RSE	PSNR
Carphone						
TRPCA-DFT	185.88	0.0658	30.5242	185.47	0.0678	30.2640
TRPCA-DCT	215.57	0.0655	30.5618	216.82	0.0673	30.3350
ScaledGD-Tucker	53.77	0.1384	24.0724	53.57	0.1364	24.1950
EAPT-DFT	30.42	0.0648	30.6570	30.28	0.0654	30.5801
EAPT-DCT	**25.62**	0.0679	30.2524	25.70	0.0685	30.1802
ScaledGD-DFT	35.09	**0.0603**	**31.2809**	33.08	**0.0624**	**30.9828**
ScaledGD-DCT	26.80	0.0632	30.8775	**25.53**	0.0655	30.5617
Coastguard						
TRPCA-DFT	165.00	**0.0511**	**31.7435**	164.15	**0.0537**	**31.3203**
TRPCA-DCT	196.24	0.0570	30.8040	199.77	0.0595	30.4273
ScaledGD-Tucker	39.37	0.1442	22.7378	39.29	0.1432	22.7957
EAPT-DFT	24.30	0.0611	30.2289	24.24	0.0621	30.0861
EAPT-DCT	23.64	0.0734	28.6131	23.54	0.0748	28.4520
ScaledGD-DFT	23.23	0.0578	30.6929	22.96	0.0597	30.4192
ScaledGD-DCT	**20.81**	0.0696	29.0768	**20.64**	0.0723	28.7491

Input   Ground Truth   TRPCA-DFT   TRPCA-DCT   ScaledGD-Tucker   EAPT-DFT   EAPT-DCT   ScaledGD-DFT   ScaledGD-DCT

Figure 5: Video denoising examples of "Carphone" (top) and "Coastguard" (bottom) for $\alpha = 0.1$.

naming convention is the same for TRPCA and EAPT. We fix $\lambda = 1/\sqrt{\max(n_1, n_2)\ell}$ for TRPCA and tune the parameters of ScaledGD-Tucker and EAPT to achieve their best performance.

Video denoising. Here, we compare the performance of different methods on video denoising. Specifically, we select two video sequences "Carphone" and "Coastguard" from YUV video sequences[1] for comparison. We randomly select α-fraction of pixels in each frame and add white Gaussian noise with variance 0.1 on these pixels. We adopt two metrics, namely, the Peak Signal-to-Noise Ratio (PSNR) and the relative standard error (RSE) for evaluation. For ScaledGD, the parameters are set to be $\zeta_0 = 0.5$, $\zeta_1 = 0.5$, $\eta = 0.5$,

---

1. http://trace.eas.asu.edu/yuv/index.html

(a) DFT   (b) DCT
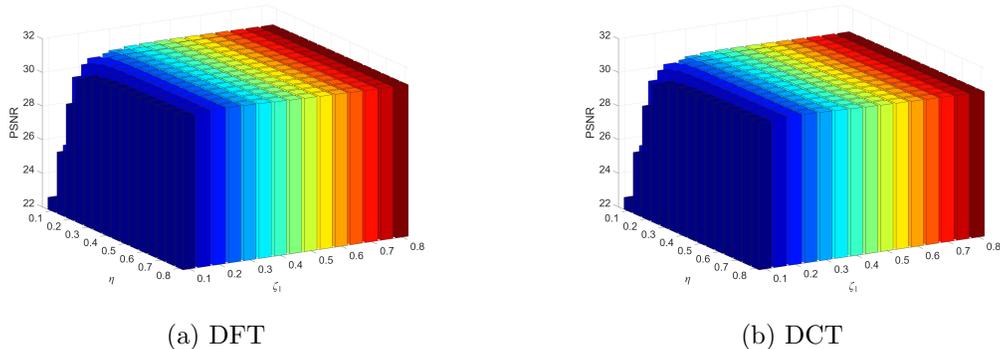
Figure 6: Parameter sensitivity analysis with respect to ζ_1 and η on the "Carphone" sequence.

$\rho = 0.9$ and $r = 20$. The numerical results in Table 1 demonstrate that t-SVD based methods achieve much better performance than ScaledGD-Tucker in terms of higher PSNR values and lower RSE values, which suggests that t-SVD provides a more realistic modelling scenario compared to Tucker decomposition for such data. ScaledGD-DFT achieves superior recovery accuracy compared to the state-of-the-art methods for "Carphone" sequence. While the PSNR value for ScaledGD-DFT is slightly lower than TRPCA-DFT for "Coastguard" sequence, the running time of TRPCA-DFT is 7 times of that of ScaledGD-DFT because ScaledGD eliminates the need for full t-SVD computations within TRPCA. Figure 5 provides two visual examples depicting the recovery results.

We also conduct experiments on "Carphone" sequence to show the effect of the parameters ζ_1 and η on the PSNR value by fixing $\zeta_0 = 0.5$ and $\rho = 0.9$. The PSNR results with different combinations of ζ_1 and η for DFT and DCT are given in Figure 6a and Figure 6b, respectively. It is evident that competitive performance can be obtained over a wide range of parameters, e.g., the PSNR value can be above 31 when $\zeta_1, \eta \geq 0.2$ for DFT.

Video background subtraction. We show the effectiveness of the proposed ScaledGD on the Background Models Challenge (BMC) dataset[2] (Vacavant et al., 2013) for video background subtraction. Here, the low-rank tensor corresponds to the relatively static background across frames, while the foreground consisting of moving objects, which are usually sparsely distributed in the video frames and can be viewed as outliers, correspond to the sparse tensor. Thus, it is reasonable to apply tensor RPCA to separate the background and foreground. The dataset contains 9 real video sequences and we use all these sequences for both qualitative and quantitative analysis. To evaluate the performance of ScaledGD, the Precision, Recall, and F-measure, are used as basic evaluation metrics. For both ScaledGD-DFT and ScaledGD-DCT, we fix $\zeta_0 = 0.15$, $\zeta_1 = 0.15$, $\eta = 0.85$, $\rho = 0.95$, $r = 5$. As can be seen from Table 2, EAPT-DCT takes the least running time and ScaledGD-DCT is only a little slower than EAPT-DCT. While ScaledGD-Tucker tends to give us a high recall value, ScaledGD-DFT and ScaledGD-DCT achieve the highest F-measure scores for 6 videos. For Video 7, the performance of ScaledGD-DFT and ScaledGD-DCT is also comparable to that

---

2. `http://backgroundmodelschallenge.eu/`

Table 2: Background subtraction results on the BMC dataset. The best result is shown in bold and the second best on F-measure and running time (in seconds) is underlined.

Methods	Metrics	Video 1	Video 2	Video 3	Video 4	Video 5	Video 6	Video 7	Video 8	Video 9
TRPCA -DFT	Precision	0.5618	**0.7175**	0.4657	**0.3893**	**0.6078**	**0.6378**	0.6238	**0.5967**	0.3094
	Recall	0.6367	0.5500	0.9147	0.7488	0.7036	0.6368	0.6088	0.6776	0.8428
	F-measure	0.5812	0.5565	**0.6119**	**0.4741**	0.5733	0.5861	<u>0.5650</u>	0.6211	0.4212
	Time	102.86	124.17	14.83	80.36	105.59	127.40	145.10	70.87	76.64
TRPCA -DCT	Precision	0.5559	0.7154	0.4646	0.3877	0.6072	0.6374	**0.6245**	0.5912	0.3160
	Recall	0.6340	0.5512	0.9143	0.7510	0.7082	0.6401	0.6107	0.6765	0.8434
	F-measure	0.5766	0.5567	<u>0.6107</u>	<u>0.4733</u>	0.5750	0.5881	**0.5671**	0.6174	0.4306
	Time	137.14	165.57	14.32	106.39	142.36	168.28	189.68	93.19	101.81
ScaledGD -Tucker	Precision	0.2466	0.4249	0.0824	0.0297	0.1972	0.2241	0.2247	0.1751	0.0156
	Recall	**0.8107**	**0.8304**	**0.9503**	**0.8908**	**0.9305**	**0.8277**	**0.8109**	**0.8167**	**0.8887**
	F-measure	0.3546	0.5356	0.1504	0.0550	0.2952	0.3244	0.3184	0.2755	0.0301
	Time	18.96	26.76	5.55	15.64	21.17	23.20	27.02	14.24	17.63
EAPT -DFT	Precision	0.5443	0.6271	0.0292	0.0887	0.4789	0.5254	0.5275	0.5138	0.1657
	Recall	0.7430	0.6423	0.3698	0.7552	0.8676	0.7200	0.6797	0.7846	0.7693
	F-measure	0.6220	0.5795	0.0537	0.1338	0.5491	0.5601	0.5436	0.6042	0.2518
	Time	12.38	15.55	3.87	10.25	13.24	14.64	16.34	9.62	11.03
EAPT -DCT	Precision	0.5603	0.6342	0.0233	0.0705	0.4918	0.5311	0.4574	0.4545	0.1096
	Recall	0.7464	0.6513	0.3992	0.7647	0.8772	0.7252	0.7210	0.7918	0.7594
	F-measure	0.6338	<u>0.5912</u>	0.0437	0.1127	0.5605	0.5701	0.5154	0.5513	0.1786
	Time	**7.37**	**9.15**	**2.30**	**5.79**	**7.78**	**8.61**	**9.22**	**5.70**	**6.51**
ScaledGD -DFT	Precision	0.6181	0.6849	**0.6467**	0.3705	0.5382	0.5831	0.5775	0.5782	0.6085
	Recall	0.7031	0.6046	0.3387	0.7209	0.8263	0.6837	0.6551	0.7580	0.7225
	F-measure	<u>0.6483</u>	0.5879	0.4087	0.4220	<u>0.5841</u>	<u>0.5891</u>	0.5636	**0.6451**	<u>0.6363</u>
	Time	10.96	14.54	2.97	10.41	12.87	13.74	16.24	9.94	10.39
ScaledGD -DCT	Precision	**0.6322**	0.6916	0.6432	0.3733	0.5674	0.5844	0.5742	0.5722	**0.6095**
	Recall	0.7101	0.6057	0.3639	0.7193	0.8535	0.6881	0.6546	0.7602	0.7357
	F-measure	**0.6605**	**0.5924**	0.4338	0.4188	**0.6172**	**0.5935**	0.5620	<u>0.6410</u>	**0.6458**
	Time	<u>8.80</u>	<u>12.18</u>	<u>2.38</u>	<u>8.32</u>	<u>10.39</u>	<u>11.67</u>	<u>14.08</u>	<u>7.54</u>	<u>8.01</u>

of TRPCA-DFT and TRPCA-DCT, but the running time of TRPCA is at least 8.5 times of that of ScaledGD. The visual results depicted in Figure 7 verify that our method is competent to extract the foreground from these videos.

We conclude by noting that the choice of transformations in these experiments is agnostic to the data. Nonetheless, selection of the optimal transformation and even learning the optimal transform is an open problem for future research.

6. Conclusions

This paper developed a scaled gradient descent (ScaledGD) algorithm for factored low-rank tensor estimation based on the t-SVD under invertible linear transforms. We rigorously establish that under standard assumptions, ScaledGD only takes $\mathcal{O}(\log(1/\epsilon))$ iterations to reach ϵ-accuracy, without the dependency on the condition number of the ground truth tensor when initialized via the spectral method. There are several future directions that are worth exploring, which we briefly discuss below.
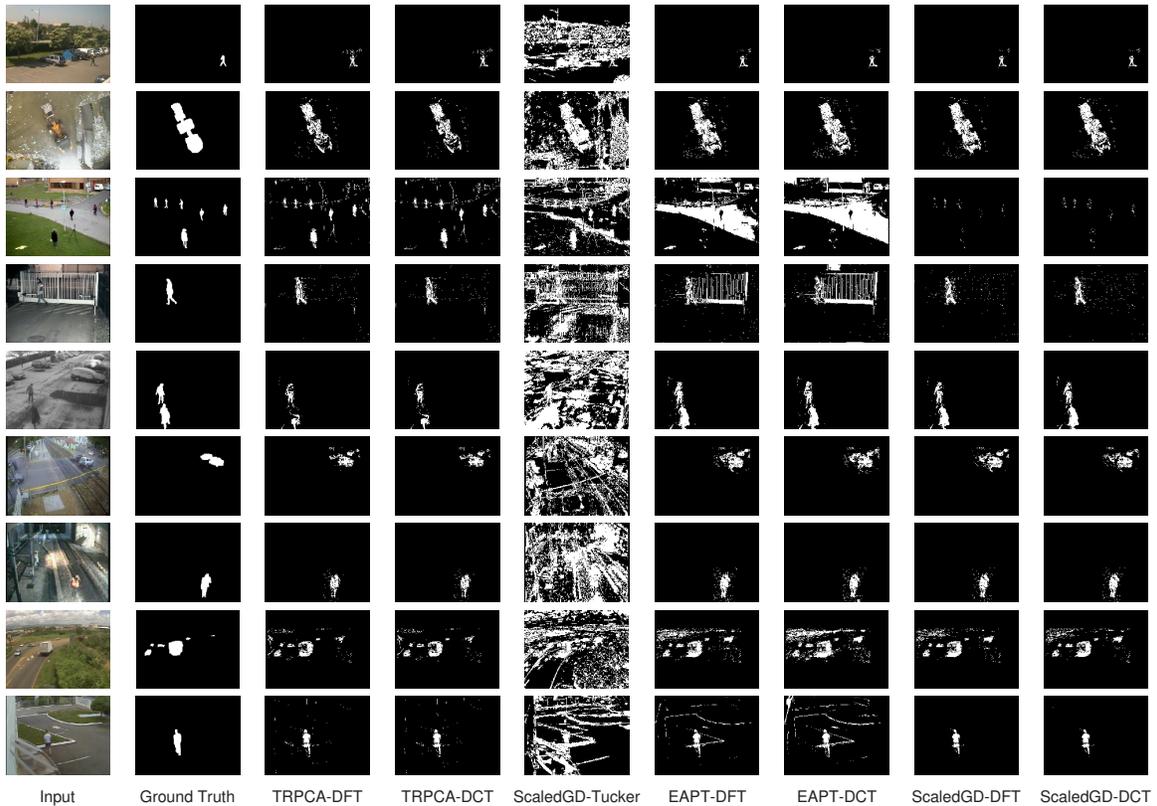
Figure 7: Video background subtraction visual results using the BMC dataset (Vacavant et al., 2013). From top to bottom are 9 sequences within the dataset.

- *Parameter estimation for tensor RPCA.* The proposed algorithm for tensor RPCA involves an iteration-varying threshold operation following a geometric decaying schedule, which contains several hyperparameters that need to be tuned carefully for real data. Our future work includes learning the optimal hyperparameters using deep unfolding and self-supervised learning (Cai et al., 2021; Dong et al., 2023a).

- *High-order extension of ScaledGD.* Based on the multilinear algebra for high-order t-SVD, it is of great interest to extend our method for high-order tensors and to unify the understanding of theoretical guarantee for low-rank tensor estimation problems when the invertible linear transforms L satisfy certain conditions.

- *Entrywise error control for tensor completion.* In this work, we aimed at minimizing the Frobenius norm of the reconstructed tensor in tensor completion. There exists another work that deals with minimizing the ℓ_∞ error for matrix completion with statistical guarantees (Ma et al., 2020). It is therefore interesting to develop similar strong entrywise error guarantees of ScaledGD for tensor completion with t-SVD.

Appendix A. Technical Lemmas

In this section, we introduce main preliminaries and useful lemmas which will be used in the proofs. We use bold calligraphic letters with arrows on top to denote tensor columns of size $n_1 \times 1 \times n_3$, e.g., $\overrightarrow{\boldsymbol{\mathcal{A}}}$. We define the $\ell_{\infty,2}$-norm and $\ell_{2,2,\infty}$-norm of a tensor $\boldsymbol{\mathcal{A}}$ as

$$\|\boldsymbol{\mathcal{A}}\|_{\infty,2} = \max\{\max_i \|\boldsymbol{\mathcal{A}}(i,:,:)\|_F, \max_j \|\boldsymbol{\mathcal{A}}(:,j,:)\|_F\},$$

and $\|\boldsymbol{\mathcal{A}}\|_{2,2,\infty} = \max_{i,k} \|\boldsymbol{\mathcal{A}}(i,:,k)\|_F$, respectively.

Definition 30 (Standard tensor basis (Lu, 2021)) *The tensor **column basis** with respect to the transform L, denoted as $\mathring{\boldsymbol{\mathfrak{e}}}_i$, is a tensor of size $n_1 \times 1 \times n_3$ with the entries of the $(i,1)$-th mode-3 tube of $L(\mathring{\boldsymbol{\mathfrak{e}}}_i)$ equaling 1 and the rest equaling 0. Similarly, the **row basis** $\mathring{\boldsymbol{\mathfrak{e}}}_j^H$ is of size $1 \times n_2 \times n_3$ with the entries of the $(1,j)$-th mode-3 tube of $L(\mathring{\boldsymbol{\mathfrak{e}}}_j^H)$ equaling to 1 and the rest equaling to 0. The **tube basis** $\dot{\boldsymbol{\mathfrak{e}}}_k$ is a tensor of size $1 \times 1 \times n_3$ with the $(1,1,k)$-th entry of $L(\dot{\boldsymbol{\mathfrak{e}}}_k)$ equaling 1 and the rest equaling 0.*

Denote $\bar{\boldsymbol{\mathfrak{e}}}_{ijk}$ as a unit tensor with only the (i,j,k)-th entry equaling 1 and others equaling 0. Based on Definition 30, $\bar{\boldsymbol{\mathfrak{e}}}_{ijk}$ can be expressed as

$$\bar{\boldsymbol{\mathfrak{e}}}_{ijk} = L(\mathring{\boldsymbol{\mathfrak{e}}}_i *_{\boldsymbol{\Phi}} \dot{\boldsymbol{\mathfrak{e}}}_k *_{\boldsymbol{\Phi}} \mathring{\boldsymbol{\mathfrak{e}}}_j^H). \tag{32}$$

Then for any tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have $\boldsymbol{\mathcal{A}}_{i,j,k} = \langle \boldsymbol{\mathcal{A}}, \bar{\boldsymbol{\mathfrak{e}}}_{ijk} \rangle$ and

$$\boldsymbol{\mathcal{A}} = \sum_{i,j,k} \langle \boldsymbol{\mathcal{A}}, \bar{\boldsymbol{\mathfrak{e}}}_{ijk} \rangle \bar{\boldsymbol{\mathfrak{e}}}_{ijk}.$$

Definition 31 *For any $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the projection onto $\boldsymbol{\Omega}$ is defined as*

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{A}}) = \sum_{ijk} \delta_{ijk} \boldsymbol{\mathcal{A}}_{i,j,k} \bar{\boldsymbol{\mathfrak{e}}}_{ijk},$$

where $\delta_{ijk} = 1_{(i,j,k) \in \boldsymbol{\Omega}}$ and $1_{(\cdot)}$ denotes the indicator function.

Definition 32 *Let $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $\mathrm{rank}_t(\boldsymbol{\mathcal{M}}) = r$ and its skinny t-SVD be $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{U}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}^H$. We denote $\boldsymbol{T}$ by the set*

$$\boldsymbol{T} = \{\boldsymbol{\mathcal{U}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Z}}^H + \boldsymbol{\mathcal{W}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}^H | \boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_2 \times r \times n_3}, \boldsymbol{\mathcal{W}} \in \mathbb{R}^{n_1 \times r \times n_3}\}, \tag{33}$$

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

The projections onto $\boldsymbol{T}$ and its complementary set $\boldsymbol{T}^{\perp}$ are respectively denoted as

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}}) = \boldsymbol{\mathcal{U}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{U}}^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}^H - \boldsymbol{\mathcal{U}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{U}}^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}^H, \tag{34}$$

and

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{T}^{\perp}}(\boldsymbol{\mathcal{A}}) = \boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}}) = (\boldsymbol{\mathcal{I}}_{n_1} - \boldsymbol{\mathcal{U}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{U}}^H) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{I}}_{n_2} - \boldsymbol{\mathcal{V}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}^H).$$

In the following, we use $\boldsymbol{\mathcal{Q}}_t$ to denote the optimal alignment tensor between $\boldsymbol{\mathcal{F}}_t = \begin{bmatrix} \boldsymbol{\mathcal{L}}_t \\ \boldsymbol{\mathcal{R}}_t \end{bmatrix}$ and $\boldsymbol{\mathcal{F}}_{\star} = \begin{bmatrix} \boldsymbol{\mathcal{L}}_{\star} \\ \boldsymbol{\mathcal{R}}_{\star} \end{bmatrix}$. For notational convenience, we denote $\boldsymbol{\mathcal{L}}_{\sharp} := \boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t$, $\boldsymbol{\mathcal{R}}_{\sharp} := \boldsymbol{\mathcal{R}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H}$, $\boldsymbol{\mathcal{L}}_{\triangle} := \boldsymbol{\mathcal{L}}_{\sharp} - \boldsymbol{\mathcal{L}}_{\star}$, $\boldsymbol{\mathcal{R}}_{\triangle} := \boldsymbol{\mathcal{R}}_{\sharp} - \boldsymbol{\mathcal{R}}_{\star}$, and $\boldsymbol{\mathcal{S}}_{\triangle} := \boldsymbol{\mathcal{S}}_{t+1} - \boldsymbol{\mathcal{S}}_{\star}$.

A.1 Tensor Algebra

Lemma 33 *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then*

$$\bar{\mathcal{X}}(:,:,k) = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_{n_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}_{k,\cdot}^T & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Phi}_{k,\cdot}^T & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Phi}_{k,\cdot}^T \end{bmatrix} := \boldsymbol{X}_{(1)} \widetilde{\boldsymbol{\Phi}}_k,$$

where each $\boldsymbol{X}_j$ is obtained by "squeezing" each sample $\mathcal{X}_{(j)} \in \mathbb{R}^{n_1 \times 1 \times n_3}$ into a matrix, i.e., $\boldsymbol{X}_j = \texttt{squeeze}(\mathcal{X}_{(j)}) \in \mathbb{R}^{n_1 \times n_3}$.

Proof The (i,j,k)-th entry of $\bar{\mathcal{X}}$ can be written as

$$\bar{\mathcal{X}}_{i,j,k} = \sum_{k'=1}^{n_3} \boldsymbol{\Phi}_{k,k'} \mathcal{X}_{i,j,k'} = \mathcal{X}(i,j,:) \boldsymbol{\Phi}_{k,\cdot}^T.$$

Then we have

$$\bar{\mathcal{X}}(:,:,k)$$

$$= \begin{bmatrix} \texttt{squeeze}(\mathcal{X}_{(1)}) & \texttt{squeeze}(\mathcal{X}_{(2)}) & \cdots & \texttt{squeeze}(\mathcal{X}_{(n_2)}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}_{k,\cdot}^T & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Phi}_{k,\cdot}^T & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Phi}_{k,\cdot}^T \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_{n_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}_{k,\cdot}^T & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Phi}_{k,\cdot}^T & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Phi}_{k,\cdot}^T \end{bmatrix} := \boldsymbol{X}_{(1)} \widetilde{\boldsymbol{\Phi}}_k,$$

where $\boldsymbol{X}_{(1)} \in \mathbb{R}^{n_1 \times n_2 n_3}$ and $\widetilde{\boldsymbol{\Phi}}_k \in \mathbb{C}^{n_2 n_3 \times n_2}$. $\blacksquare$

Lemma 34 *Let $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is α-sparse, then*

$$\|\mathcal{S}\| \le \frac{\alpha\sqrt{\ell}}{2}(n_1 + n_2 n_3)\|\mathcal{S}\|_\infty \quad \text{and} \quad \|\mathcal{S}\|_{2,\infty} \le \sqrt{\alpha n_2 n_3}\|\mathcal{S}\|_\infty.$$

Proof We only need to prove the first claim, since the second claim can be directly followed by the fact that $\mathcal{S}$ has at most α-fraction nonzero entries along each mode-2 tube of $\mathcal{S}$. Let

$\boldsymbol{x} \in \mathbb{R}^{n_1}$ and $\boldsymbol{z} \in \mathbb{R}^{n_2 n_3}$ be unit vectors, using $ab \leq (a^2 + b^2)/2$, we have

$$\|\boldsymbol{\mathcal{S}}\| = \|\bar{\boldsymbol{\mathcal{S}}}\| = \max_{k=1,\ldots,n_3} \|\bar{\boldsymbol{\mathcal{S}}}(:,:,k)\| = \max_{k=1,\ldots,n_3} \|\boldsymbol{S}_{(1)}\widetilde{\boldsymbol{\Phi}}_k\| \leq \max_{k=1,\ldots,n_3} \|\boldsymbol{S}_{(1)}\|\|\widetilde{\boldsymbol{\Phi}}_k\|$$

$$= \max_{k=1,\ldots,n_3} \sqrt{\ell}\|\boldsymbol{S}_{(1)}\| = \sqrt{\ell} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} x_i \boldsymbol{\mathcal{S}}_{i,j,k} z_{(j-1)n_3+k}$$

$$\leq \frac{\sqrt{\ell}}{2} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} (x_i^2 + z_{(j-1)n_3+k}^2)\boldsymbol{\mathcal{S}}_{i,j,k}$$

$$\leq \frac{\sqrt{\ell}}{2}(\alpha n_2 n_3 + \alpha n_1)\|\boldsymbol{\mathcal{S}}\|_\infty = \frac{\alpha\sqrt{\ell}}{2}(n_1 + n_2 n_3)\|\boldsymbol{\mathcal{S}}\|_\infty,$$

where the last inequality follows from the assumption that $\boldsymbol{\mathcal{S}} \in \mathscr{S}_\alpha$. ∎

Lemma 35 *Let $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{n_4 \times n_2 \times n_3}$ be two tensors, then*

$$\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}^H\|_\infty \leq \sqrt{\ell}\|\boldsymbol{\mathcal{A}}\|_{2,\infty}\|\boldsymbol{\mathcal{B}}\|_{2,\infty}.$$

Proof By the definition of the transform based t-product, we have

$$\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}^H\|_\infty = \max_{i,i'} \|\sum_{j=1}^{n_2} \boldsymbol{\mathcal{A}}(i,j,:) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}(i',j,:)\|_\infty \leq \max_{i,i'} \|\sum_{j=1}^{n_2} \boldsymbol{\mathcal{A}}(i,j,:) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}(i',j,:)\|_2$$

$$\leq \max_{i,i'} \sum_{j=1}^{n_2} \|\boldsymbol{\mathcal{A}}(i,j,:) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}(i',j,:)\|_2 = \max_{i,i'} \sum_{j=1}^{n_2} \frac{1}{\sqrt{\ell}}\|\bar{\boldsymbol{\mathcal{A}}}(i,j,:)\triangle\bar{\boldsymbol{\mathcal{B}}}(i',j,:)\|_2$$

$$\leq \max_{i,i'} \sum_{j=1}^{n_2} \frac{1}{\sqrt{\ell}}\|\bar{\boldsymbol{\mathcal{A}}}(i,j,:)\|_2\|\bar{\boldsymbol{\mathcal{B}}}(i',j,:)\|_2 = \max_{i,i'} \sum_{j=1}^{n_2} \sqrt{\ell}\|\boldsymbol{\mathcal{A}}(i,j,:)\|_2\|\boldsymbol{\mathcal{B}}(i',j,:)\|_2$$

$$\leq \max_{i,i'} \sqrt{\ell}\|\boldsymbol{\mathcal{A}}(i,:,:)\|_F\|\boldsymbol{\mathcal{B}}(i',:,:)\|_F \leq \sqrt{\ell}\|\boldsymbol{\mathcal{A}}\|_{2,\infty}\|\boldsymbol{\mathcal{B}}\|_{2,\infty}.$$

∎

Lemma 36 *Let $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ be two tensors. Assume the multi-rank of $\boldsymbol{\mathcal{B}}$ is $\boldsymbol{r}$ and let $s_r = \sum_{k=1}^{n_3} r_k$, then*

$$\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_F \geq \|\boldsymbol{\mathcal{A}}\|_F \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{B}}) \quad \text{and} \quad \|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_F \leq \|\boldsymbol{\mathcal{A}}\|_F \|\boldsymbol{\mathcal{B}}\|.$$

Moreover,

$$\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_{2,\infty} \geq \|\boldsymbol{\mathcal{A}}\|_{2,\infty} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{B}}) \quad \text{and} \quad \|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_{2,\infty} \leq \|\boldsymbol{\mathcal{A}}\|_{2,\infty} \|\boldsymbol{\mathcal{B}}\|.$$

Proof To prove the first claim, we start with the definition of Frobenius norm

$$\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_F^2 = \frac{1}{\ell}\|\bar{\boldsymbol{\mathcal{A}}}\triangle\bar{\boldsymbol{\mathcal{B}}}\|_F^2 = \frac{1}{\ell}\|\bar{\boldsymbol{A}}\bar{\boldsymbol{B}}\|_F^2 = \frac{1}{\ell}\text{tr}(\bar{\boldsymbol{A}}\bar{\boldsymbol{B}}\bar{\boldsymbol{B}}^H\bar{\boldsymbol{A}}^H) = \frac{1}{\ell}\text{tr}(\bar{\boldsymbol{B}}\bar{\boldsymbol{B}}^H\bar{\boldsymbol{A}}^H\bar{\boldsymbol{A}}).$$

By using the inequalities for matrix trace, we have

$$\lambda_{\min}(\bar{\boldsymbol{B}}\bar{\boldsymbol{B}}^H)\mathrm{tr}(\bar{\boldsymbol{A}}^H\bar{\boldsymbol{A}}) \leq \mathrm{tr}(\bar{\boldsymbol{B}}\bar{\boldsymbol{B}}^H\bar{\boldsymbol{A}}^H\bar{\boldsymbol{A}}) \leq \lambda_{\max}(\bar{\boldsymbol{B}}\bar{\boldsymbol{B}}^H)\mathrm{tr}(\bar{\boldsymbol{A}}^H\bar{\boldsymbol{A}}),$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum and maximum eigenvalue, respectively. This implies

$$\sigma_{\min}^2(\bar{\boldsymbol{B}})\mathrm{tr}(\bar{\boldsymbol{A}}^H\bar{\boldsymbol{A}}) \leq \ell\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_F^2 \leq \sigma_{\max}^2(\bar{\boldsymbol{B}})\mathrm{tr}(\bar{\boldsymbol{A}}^H\bar{\boldsymbol{A}}),$$

where $\sigma_{\min}$ and $\sigma_{\max}$ denote the minimum and maximum singular value, respectively. Hence, we have

$$\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_F^2 \geq \sigma_{\min}^2(\bar{\boldsymbol{B}})\frac{1}{\ell}\mathrm{tr}(\bar{\boldsymbol{A}}^H\bar{\boldsymbol{A}}) = \bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{B}})\|\boldsymbol{\mathcal{A}}\|_F^2$$

$$\text{and} \quad \|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_F^2 \leq \sigma_{\max}^2(\bar{\boldsymbol{B}})\frac{1}{\ell}\mathrm{tr}(\bar{\boldsymbol{A}}^H\bar{\boldsymbol{A}}) = \|\boldsymbol{\mathcal{B}}\|^2\|\boldsymbol{\mathcal{A}}\|_F^2.$$

Taking the square root on both sides to arrive at the first claim. The second conclusion is an easy consequence of the first one as

$$\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_{2,\infty} = \max_i \|\boldsymbol{\mathcal{A}}(i,:,:) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_F \geq \max_i \|\boldsymbol{\mathcal{A}}(i,:,:)\|_F\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{B}}) = \|\boldsymbol{\mathcal{A}}\|_{2,\infty}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{B}})$$

$$\text{and} \quad \|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_{2,\infty} = \max_i \|\boldsymbol{\mathcal{A}}(i,:,:) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_F \leq \max_i \|\boldsymbol{\mathcal{A}}(i,:,:)\|_F\|\boldsymbol{\mathcal{B}}\| = \|\boldsymbol{\mathcal{A}}\|_{2,\infty}\|\boldsymbol{\mathcal{B}}\|.$$

∎

Lemma 37 *Let* $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *and* $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ *be two tensors, then*

$$\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_{2,\infty} \leq \sqrt{n_2\ell}\|\boldsymbol{\mathcal{A}}\|_{2,\infty}\|\boldsymbol{\mathcal{B}}\|_{2,\infty}.$$

Proof Based on the definition of the $\ell_{2,\infty}$-norm, we have

$$\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_{2,\infty} = \frac{1}{\sqrt{\ell}}\|\bar{\boldsymbol{\mathcal{A}}}\triangle\bar{\boldsymbol{\mathcal{B}}}\|_{2,\infty} = \max_i \frac{1}{\sqrt{\ell}}\|\bar{\boldsymbol{\mathcal{A}}}(i,:,:)\triangle\bar{\boldsymbol{\mathcal{B}}}\|_F.$$

Note that

$$\|\bar{\boldsymbol{\mathcal{A}}}(i,:,:)\triangle\bar{\boldsymbol{\mathcal{B}}}\|_F = \sqrt{\sum_{k=1}^{n_3}\|\bar{\boldsymbol{\mathcal{A}}}(i,:,k)\bar{\boldsymbol{\mathcal{B}}}(:,:,k)\|_2^2} = \sqrt{\sum_{k=1}^{n_3}\sum_{j=1}^{n_4}(\langle\bar{\boldsymbol{\mathcal{A}}}(i,:,k),\bar{\boldsymbol{\mathcal{B}}}(:,j,k)\rangle)^2}$$

$$\leq \sqrt{\sum_{k=1}^{n_3}\sum_{j=1}^{n_4}\|\bar{\boldsymbol{\mathcal{A}}}(i,:,k)\|_2^2\|\bar{\boldsymbol{\mathcal{B}}}(:,j,k)\|_2^2} = \sqrt{\sum_{k=1}^{n_3}\|\bar{\boldsymbol{\mathcal{A}}}(i,:,k)\|_2^2\|\bar{\boldsymbol{\mathcal{B}}}(:,:,k)\|_F^2}$$

$$\leq \|\bar{\boldsymbol{\mathcal{A}}}(i,:,:)\|_F\|\bar{\boldsymbol{\mathcal{B}}}\|_F = \ell\|\boldsymbol{\mathcal{A}}(i,:,:)\|_F\|\boldsymbol{\mathcal{B}}\|_F.$$

Thus,

$$\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_{2,\infty} \leq \max_i \sqrt{\ell}\|\boldsymbol{\mathcal{A}}(i,:,:)\|_F\|\boldsymbol{\mathcal{B}}\|_F = \sqrt{\ell}\|\boldsymbol{\mathcal{A}}\|_{2,\infty}\|\boldsymbol{\mathcal{B}}\|_F \leq \sqrt{n_2\ell}\|\boldsymbol{\mathcal{A}}\|_{2,\infty}\|\boldsymbol{\mathcal{B}}\|_{2,\infty}.$$

∎

Lemma 38 *Let L be any invertible linear transform in* (6) *and it satisfies* (5). *For any $i \in [n_1]$, $j \in [n_2]$ and $k \in [n_3]$, given $\bar{\mathbf{e}}_{ijk}$ in* (32), *we have the following properties*

$$\|\bar{\mathbf{e}}_{ijk}\| \leq \sqrt{\ell}, \tag{35}$$

$$\left\| \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} (\bar{\mathbf{e}}_{ijk}^H *_{\mathbf{\Phi}} \bar{\mathbf{e}}_{ijk}) \right\| \leq n_1 \ell, \tag{36}$$

and

$$\left\| \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} (\bar{\mathbf{e}}_{ijk} *_{\mathbf{\Phi}} \bar{\mathbf{e}}_{ijk}^H) \right\| \leq n_2 \ell. \tag{37}$$

Proof To simplify the notation, we denote $\boldsymbol{\mathcal{A}} = \bar{\mathbf{e}}_{ijk}$ in this proof. By the definition of the tensor spectral norm, we have

$$\|\boldsymbol{\mathcal{A}}\| = \|\bar{\boldsymbol{A}}\| = \max_{k'=1,\dots,n_3} \|\bar{\boldsymbol{A}}^{(k')}\|.$$

Since $\boldsymbol{\mathcal{A}}$ is the unit tensor, the (i,j)-th mode-3 tube of $\bar{\boldsymbol{A}} = L(\bar{\mathbf{e}}_{ijk})$ is the only nonzero mode-3 tube and its entries are the same as the k-th column of $\mathbf{\Phi}$, i.e., $\bar{\boldsymbol{A}}_{i,j,k'} = \mathbf{\Phi}_{k',k}$. Then $\|\bar{\boldsymbol{A}}^{(k')}\| = |\mathbf{\Phi}_{k',k}|$ and we have

$$\|\boldsymbol{\mathcal{A}}\| = \max_{k'=1,\dots,n_3} \|\bar{\boldsymbol{A}}^{(k')}\| = \max_{k'=1,\dots,n_3} |\mathbf{\Phi}_{k',k}| \leq \sqrt{\ell},$$

where the inequality comes from $|\mathbf{\Phi}_{k',k}| = \sqrt{|\mathbf{\Phi}_{k',k}|^2} \leq \sqrt{\sum_{k'=1}^{n_3} |\mathbf{\Phi}_{k',k}|^2} = \sqrt{\ell}$. Therefore, (35) is verified. To prove (36), let $\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{A}}^H *_{\mathbf{\Phi}} \boldsymbol{\mathcal{A}}$, then $\bar{\boldsymbol{\mathcal{B}}} = \bar{\boldsymbol{\mathcal{A}}}^H \triangle \bar{\boldsymbol{\mathcal{A}}}$, which means that the (j,j)-th mode-3 tube of $\bar{\boldsymbol{\mathcal{B}}}$ is the only nonzero mode-3 tube and its k'-th entry is $\bar{\boldsymbol{\mathcal{B}}}_{j,j,k'} = |\mathbf{\Phi}_{k',k}|^2$. Thus the (j,j)-th mode-3 tube of

$$\sum_{k=1}^{n_3} L(\bar{\mathbf{e}}_{ijk}^H *_{\mathbf{\Phi}} \bar{\mathbf{e}}_{ijk}) = \sum_{k=1}^{n_3} \bar{\boldsymbol{\mathcal{B}}}$$

is the only nonzero mode-3 tube and all of its entries equal ℓ. Further,

$$\sum_{j=1}^{n_2} \sum_{k=1}^{n_3} L(\bar{\mathbf{e}}_{ijk}^H *_{\mathbf{\Phi}} \bar{\mathbf{e}}_{ijk}) = \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \bar{\boldsymbol{\mathcal{B}}}$$

is an f-diagonal tensor and all the entries on the diagonal equal ℓ. Finally,

$$\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} L(\bar{\mathbf{e}}_{ijk}^H *_{\mathbf{\Phi}} \bar{\mathbf{e}}_{ijk}) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \bar{\boldsymbol{\mathcal{B}}}$$

is also f-diagonal and all of its entries on the (j,j)-th mode-3 tube equal $n_1 \ell$, $j \in [n_2]$. It is obvious that each frontal slice of such a tensor has the spectral norm $n_1 \ell$. Therefore, we have (36). We can prove (37) in a similar way. ∎

Lemma 39 *Given* $\mathcal{P}_T$ *in* (34) *and assume that the tensor incoherence conditions* (10) *hold. We have*

$$\|\mathcal{P}_T(\bar{\mathbf{e}}_{ijk})\|_F^2 \leq \frac{\mu s_r(n_1 + n_2)}{n_1 n_2 n_3}.$$

Proof Note that $\mathcal{P}_T$ is self-adjoint. So we have

$$\begin{aligned}
&\|\mathcal{P}_T(\bar{\mathbf{e}}_{ijk})\|_F^2 \\
=&\langle \mathcal{P}_T(\bar{\mathbf{e}}_{ijk}), \bar{\mathbf{e}}_{ijk}\rangle \\
=&\langle \mathcal{U} *_\Phi \mathcal{U}^H *_\Phi \bar{\mathbf{e}}_{ijk} + \bar{\mathbf{e}}_{ijk} *_\Phi \mathcal{V} *_\Phi \mathcal{V}^H, \bar{\mathbf{e}}_{ijk}\rangle + \langle \mathcal{U} *_\Phi \mathcal{U}^H *_\Phi \bar{\mathbf{e}}_{ijk} *_\Phi \mathcal{V} *_\Phi \mathcal{V}^H, \bar{\mathbf{e}}_{ijk}\rangle.
\end{aligned}$$

Note that $\bar{\mathbf{e}}_{ijk}$ is the unit tensor with the (i, j, k)-th entry equaling to 1 and the rest equaling to 0. Hence, the (i, j)-th mode-3 tube of $\bar{\mathbf{e}}_{ijk}$ equals $L(\dot{\mathbf{e}}_k)$ and the (i, j)-th mode-3 tube of $L(\bar{\mathbf{e}}_{ijk})$ equals $L(L(\dot{\mathbf{e}}_k))$. Then the only nonzero lateral slice of $L(\bar{\mathbf{e}}_{ijk})$ is the j-th lateral slice and it is equal to $L(\mathring{\mathbf{e}}_i)\triangle L(L(\dot{\mathbf{e}}_k))$. This implies that

$$\begin{aligned}
\langle \mathcal{U} *_\Phi \mathcal{U}^H *_\Phi \bar{\mathbf{e}}_{ijk}, \bar{\mathbf{e}}_{ijk}\rangle =&\frac{1}{\ell}\|\bar{\mathcal{U}}^H \triangle L(\bar{\mathbf{e}}_{ijk})\|_F^2 = \frac{1}{\ell}\|\bar{\mathcal{U}}^H \triangle L(\mathring{\mathbf{e}}_i)\triangle L(L(\dot{\mathbf{e}}_k))\|_F^2 \\
=&\|\mathcal{U}^H *_\Phi \mathring{\mathbf{e}}_i *_\Phi L(\dot{\mathbf{e}}_k)\|_F^2.
\end{aligned}$$

In the meanwhile, we have

$$\begin{aligned}
\|\mathcal{U}^H *_\Phi \mathring{\mathbf{e}}_i *_\Phi L(\dot{\mathbf{e}}_k)\|_F =& \frac{1}{\sqrt{\ell}}\|L(\mathcal{U}^H)\triangle L(\mathring{\mathbf{e}}_i)\triangle L(L(\dot{\mathbf{e}}_k))\|_F \leq \frac{1}{\sqrt{\ell}}\|L(\mathcal{U}^H)\triangle L(\mathring{\mathbf{e}}_i)\|_F\|L(L(\dot{\mathbf{e}}_k))\|_F \\
=& \sqrt{\ell}\|\mathcal{U}^H *_\Phi \mathring{\mathbf{e}}_i\|_F \leq \sqrt{\frac{\mu s_r}{n_1 n_3}},
\end{aligned}$$

where we use $\|L(L(\dot{\mathbf{e}}_k))\|_F = \sqrt{\ell}\|L(\dot{\mathbf{e}}_k)\|_F = \sqrt{\ell}$. Similarly, we can also have $\|\mathcal{V}^H *_\Phi \mathring{\mathbf{e}}_j *_\Phi L(\dot{\mathbf{e}}_k)\|_F \leq \sqrt{\frac{\mu s_r}{n_2 n_3}}$. Therefore, we have

$$\begin{aligned}
\|\mathcal{P}_T(\bar{\mathbf{e}}_{ijk})\|_F^2 =&\|\mathcal{U}^H *_\Phi \mathring{\mathbf{e}}_i *_\Phi L(\dot{\mathbf{e}}_k)\|_F^2 + \|\mathcal{V}^H *_\Phi \mathring{\mathbf{e}}_j *_\Phi L(\dot{\mathbf{e}}_k)\|_F^2 - \|\mathcal{U}^H *_\Phi \bar{\mathbf{e}}_{ijk} *_\Phi \mathcal{V}\|_F^2 \\
\leq&\|\mathcal{U}^H *_\Phi \mathring{\mathbf{e}}_i *_\Phi L(\dot{\mathbf{e}}_k)\|_F^2 + \|\mathcal{V}^H *_\Phi \mathring{\mathbf{e}}_j *_\Phi L(\dot{\mathbf{e}}_k)\|_F^2 \\
\leq&\frac{\mu s_r(n_1 + n_2)}{n_1 n_2 n_3}.
\end{aligned}$$

The proof is completed. ∎

Using the same proof technique in Lu et al. (2020, Lemma 4.2), we have the following result.

Lemma 40 *Suppose* $\Omega \sim \text{Ber}(p)$, *and* T *is defined in* (33). *Then with high probability,*

$$\left\|\mathcal{P}_T - \frac{1}{p}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T\right\| \leq \epsilon,$$

provided that $p \geq c\epsilon^{-2}\mu s_r(n_1 + n_2)\log((n_1 \vee n_2)n_3)/(n_1 n_2 n_3)$ *for some numerical constant* $c > 0$.

A.2 Distance Metric

Lemma 41 *Fix any factor tensor* $\boldsymbol{\mathcal{F}} = \begin{bmatrix} \boldsymbol{\mathcal{L}} \\ \boldsymbol{\mathcal{R}} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2) \times r \times n_3}$. *Suppose that*

$$\text{dist}(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}_\star) < \frac{1}{\sqrt{\ell}} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star), \tag{38}$$

then the minimizer of the above minimization problem is attained at some $\boldsymbol{\mathcal{Q}} \in \text{GL}(r)$, *i.e., the optimal alignment tensor* $\boldsymbol{\mathcal{Q}}$ *between* $\boldsymbol{\mathcal{F}}$ *and* $\boldsymbol{\mathcal{F}}_\star$ *exists.*

Proof Based on the definition of infimum and condition (38), there must exist a tensor $\widetilde{\boldsymbol{\mathcal{Q}}} \in \text{GL}(r)$ such that

$$\sqrt{\|(\bar{\boldsymbol{L}}\bar{\widetilde{\boldsymbol{Q}}} - \bar{\boldsymbol{L}}_\star)\bar{\boldsymbol{G}}_\star^{-\frac{1}{2}} \bar{\boldsymbol{G}}_\star\|_F^2 + \|(\bar{\boldsymbol{R}}\bar{\widetilde{\boldsymbol{Q}}}^{-H} - \bar{\boldsymbol{R}}_\star)\bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}\bar{\boldsymbol{G}}_\star\|_F^2}$$
$$= \sqrt{\ell}\sqrt{\|(\boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \widetilde{\boldsymbol{\mathcal{Q}}} - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \|(\boldsymbol{\mathcal{R}} *_{\boldsymbol{\Phi}} \widetilde{\boldsymbol{\mathcal{Q}}}^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2}$$
$$\leq \epsilon \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$$

together with the relation $\|\boldsymbol{AB}\|_F \geq \|\boldsymbol{A}\|_F \sigma_{\min}(\boldsymbol{B})$ tells that

$$\sqrt{\|(\bar{\boldsymbol{L}}\bar{\widetilde{\boldsymbol{Q}}} - \bar{\boldsymbol{L}}_\star)\bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}\|_F^2 + \|(\bar{\boldsymbol{R}}\bar{\widetilde{\boldsymbol{Q}}}^{-H} - \bar{\boldsymbol{R}}_\star)\bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}\|_F^2}\, \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) \leq \epsilon \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$$

for some ϵ obeying $0 < \epsilon < 1$. It further implies that

$$\|(\bar{\boldsymbol{L}}\bar{\widetilde{\boldsymbol{Q}}} - \bar{\boldsymbol{L}}_\star)\bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}\| \vee \|(\bar{\boldsymbol{R}}\bar{\widetilde{\boldsymbol{Q}}}^{-H} - \bar{\boldsymbol{R}}_\star)\bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}\| \leq \epsilon.$$

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

Lemma 42 *For any factor tensor* $\boldsymbol{\mathcal{F}} = \begin{bmatrix} \boldsymbol{\mathcal{L}} \\ \boldsymbol{\mathcal{R}} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2) \times r \times n_3}$, *suppose that the optimal alignment tensor*

$$\boldsymbol{\mathcal{Q}} = \arg\min_{\boldsymbol{\mathcal{Q}} \in \text{GL}(r)} \|(\boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}} - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \|(\boldsymbol{\mathcal{R}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 \tag{39}$$

between $\boldsymbol{\mathcal{F}}$ *and* $\boldsymbol{\mathcal{F}}_\star$ *exists, then* $\boldsymbol{\mathcal{Q}}$ *obeys*

$$(\boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}})^H *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}} - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star = \boldsymbol{\mathcal{G}}_\star *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}^{-H} - \boldsymbol{\mathcal{R}}_\star)^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}^{-H}. \tag{40}$$

Proof Check the gradient of the objective function in (39) with respect to $\boldsymbol{\mathcal{Q}}$ and set it to zero yields

$$2\boldsymbol{\mathcal{L}}^H *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}} - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star - 2\boldsymbol{\mathcal{Q}}^{-H} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}^{-H} - \boldsymbol{\mathcal{R}}_\star)^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}^{-H} = \boldsymbol{0},$$

which implies the optimal alignment criterion (40). ∎

Lastly, following the proof in Tong et al. (2021a, Lemma 24), we connect the proposed distance to the Frobenius norm in Lemma 43.

Lemma 43 *For any factor tensor* $\boldsymbol{\mathcal{F}} = \begin{bmatrix} \boldsymbol{\mathcal{L}} \\ \boldsymbol{\mathcal{R}} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2) \times r \times n_3}$, *the distance between* $\boldsymbol{\mathcal{F}}$ *and* $\boldsymbol{\mathcal{F}}_\star$ *satisfies*

$$\mathrm{dist}(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{F}}_\star) \leq \left(\sqrt{2}+1\right)^{\frac{1}{2}} \|\boldsymbol{\mathcal{L}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}^H - \boldsymbol{\mathcal{X}}_\star\|_F.$$

A.3 Tensor Perturbation Bounds

Lemma 44 *For any* $\boldsymbol{\mathcal{L}}_\sharp \in \mathbb{R}^{n_1 \times r \times n_3}$, $\boldsymbol{\mathcal{R}}_\sharp \in \mathbb{R}^{n_2 \times r \times n_3}$, *denote* $\boldsymbol{\mathcal{L}}_\triangle := \boldsymbol{\mathcal{L}}_\sharp - \boldsymbol{\mathcal{L}}_\star$ *and* $\boldsymbol{\mathcal{R}}_\triangle := \boldsymbol{\mathcal{R}}_\sharp - \boldsymbol{\mathcal{R}}_\star$. *Suppose that* $\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\| \vee \|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\| < 1$, *then*

$$\|\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\| \leq \frac{1}{1 - \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|}; \tag{41}$$

$$\|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\| \leq \frac{1}{1 - \|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|}; \tag{42}$$

$$\|\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} - \boldsymbol{\mathcal{U}}_\star\| \leq \frac{\sqrt{2}\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|}{1 - \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|}; \tag{43}$$

$$\|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} - \boldsymbol{\mathcal{V}}_\star\| \leq \frac{\sqrt{2}\|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|}{1 - \|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|}. \tag{44}$$

Proof We only prove (41) and (43), since the claims (42) and (44) on the factor $\boldsymbol{\mathcal{R}}$ can be proved in a similar way. We first notice that

$$\|\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\| = \|\bar{\boldsymbol{L}}_\sharp (\bar{\boldsymbol{L}}_\sharp^H \bar{\boldsymbol{L}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\| = \frac{1}{\bar{\sigma}_{s_r}(\bar{\boldsymbol{L}}_\sharp \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}})}.$$

We invoke Weyl's inequality $|\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{A}}) - \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{B}})| = |\sigma_{s_r}(\bar{\boldsymbol{A}}) - \sigma_{s_r}(\bar{\boldsymbol{B}})| \leq \|\bar{\boldsymbol{A}} - \bar{\boldsymbol{B}}\|$, and use the fact that $\bar{\boldsymbol{U}}_\star = \bar{\boldsymbol{L}}_\star \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}$ satisfies $\sigma_{s_r}(\bar{\boldsymbol{U}}_\star) = 1$ to obtain

$$\sigma_{s_r}(\bar{\boldsymbol{L}}_\sharp \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}) \geq \sigma_{s_r}(\bar{\boldsymbol{L}}_\star \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}) - \|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}\| = 1 - \|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}\| = 1 - \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|.$$

Then (41) follows immediately by combining the preceding two relations. In order to prove (43), using the facts that $\boldsymbol{\mathcal{L}}_\star^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{U}}_\star = \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}$ and $(\boldsymbol{\mathcal{I}}_{n_1} - \boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp^H) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp = \boldsymbol{0}$,

we have the following decomposition

$$
\begin{aligned}
&\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} - \boldsymbol{\mathcal{U}}_\star \\
=&\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\star^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{U}}_\star - \boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}} \\
=&-\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp - \boldsymbol{\mathcal{L}}_\star)^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{U}}_\star \\
&\qquad\qquad - (\boldsymbol{\mathcal{I}}_{n_1} - \boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp^H) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}} \\
=&-\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\triangle^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{U}}_\star \\
&\qquad\qquad + (\boldsymbol{\mathcal{I}}_{n_1} - \boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp^H) *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}} \\
=&-\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\triangle^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{U}}_\star \\
&\qquad\qquad + (\boldsymbol{\mathcal{I}}_{n_1} - \boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp^H) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}.
\end{aligned}
$$

In the meanwhile, we can verify that $\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\triangle^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{U}}_\star$ and $(\boldsymbol{\mathcal{I}}_{n_1} - \boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp^H) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}$ are orthogonal, thus

$$
\begin{aligned}
&\|\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} - \boldsymbol{\mathcal{U}}_\star\|^2 \\
\leq&\|\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\triangle^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{U}}_\star\|^2 \\
&\qquad\qquad + \|(\boldsymbol{\mathcal{I}}_{n_1} - \boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp^H) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|^2 \\
\leq&\|\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|^2 \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|^2 \\
&\qquad\qquad + \|\boldsymbol{\mathcal{I}}_{n_1} - \boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp^H\|^2 \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|^2 \\
\leq&\frac{\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|^2}{(1 - \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|)^2} + \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|^2 \\
\leq&\frac{2\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|^2}{(1 - \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|)^2},
\end{aligned}
$$

where we have used (41) and the fact that $\|\boldsymbol{\mathcal{I}}_{n_1} - \boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{L}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp^H\| \leq 1$. ∎

Lemma 45 *For any $\boldsymbol{\mathcal{L}}_\sharp \in \mathbb{R}^{n_1 \times r \times n_3}$, $\boldsymbol{\mathcal{R}}_\sharp \in \mathbb{R}^{n_2 \times r \times n_3}$, denote $\boldsymbol{\mathcal{L}}_\triangle := \boldsymbol{\mathcal{L}}_\sharp - \boldsymbol{\mathcal{L}}_\star$ and $\boldsymbol{\mathcal{R}}_\triangle := \boldsymbol{\mathcal{R}}_\sharp - \boldsymbol{\mathcal{R}}_\star$, then*

$$
\begin{aligned}
\|\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{\mathcal{X}}_\star\|_F &\leq \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\star^H\|_F + \|\boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\triangle^H\|_F + \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\triangle^H\|_F \\
&\leq \left(1 + \frac{1}{2}(\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\| \vee \|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|)\right) \\
&\qquad\qquad\qquad \left(\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F + \|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F\right).
\end{aligned}
$$

Proof Using the decomposition that $\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star = \mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H + \mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H + \mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H$ and the facts that

$$\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H\|_F = \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \mathcal{V}_\star^H\|_F = \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F,$$

$$\text{and}\quad \|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F = \|\mathcal{U}_\star *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F = \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F,$$

as well as the triangle inequality, we have

$$\|\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star\|_F \le \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H\|_F + \|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F + \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F$$
$$= \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F + \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F + \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F.$$

This together with the following upper bound

$$\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F = \frac{1}{2}\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} (\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}})^H\|_F + \frac{1}{2}\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}} *_{\boldsymbol{\Phi}} (\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}})^H\|_F$$
$$= \frac{1}{2\sqrt{\ell}}\|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}(\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}})^H\|_F + \frac{1}{2\sqrt{\ell}}\|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}(\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}})^H\|_F$$
$$\le \frac{1}{2\sqrt{\ell}}\|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F\|\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}\| + \frac{1}{2\sqrt{\ell}}\|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}\|\|\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F$$
$$\le \frac{1}{2\sqrt{\ell}}(\|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}\| \vee \|\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star^{-\frac{1}{2}}\|)\Big(\|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F + \|\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F\Big)$$
$$= \frac{1}{2}(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\| \vee \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\|)\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F + \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big)$$

finishes the proof. ∎

A.4 Partial Frobenius Norm

We introduce the partial Frobenius norm

$$\|\mathcal{X}\|_{F,r} := \frac{1}{\sqrt{\ell}}\sqrt{\sum_{k=1}^{n_3}\sum_{i=1}^{r} \bar{\mathcal{G}}_{i,i,k}^2} = \|\mathbb{P}_r(\mathcal{X})\|_F \tag{45}$$

as the Frobenius norm of the tubal rank-r approximation $\mathbb{P}_r(\mathcal{X})$ defined in (9). According to Lemma 28 in Tong et al. (2021a), it is easy to verify that $\|\cdot\|_{F,r}$ is a norm, and we have the following lemma that provides useful characterizations of the partial Frobenius norm on tensors.

Lemma 46 *For any $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have*

$$\|\mathcal{X}\|_{F,r} = \max_{\widetilde{\mathcal{V}} \in \mathbb{R}^{n_2 \times r \times n_3}: \widetilde{\mathcal{V}}^H *_{\boldsymbol{\Phi}} \widetilde{\mathcal{V}} = \mathcal{I}_r} \|\mathcal{X} *_{\boldsymbol{\Phi}} \widetilde{\mathcal{V}}\|_F \tag{46}$$

$$= \max_{\widetilde{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}: \|\widetilde{\mathcal{X}}\|_F \le 1, \mathrm{rank}_t(\widetilde{\mathcal{X}}) \le r} |\langle \mathcal{X}, \widetilde{\mathcal{X}}\rangle| \tag{47}$$

$$= \max_{\widetilde{\mathcal{R}} \in \mathbb{R}^{n_2 \times r \times n_3}: \|\widetilde{\mathcal{R}}\| \le 1} \|\mathcal{X} *_{\boldsymbol{\Phi}} \widetilde{\mathcal{R}}\|_F. \tag{48}$$

Recall that $\mathbb{P}_r(\boldsymbol{X})$ denotes the best rank-r approximation of $\boldsymbol{X}$ under the Frobenius norm, see Zhang et al. (2014, Theorem 2.3.1) and Kilmer et al. (2021, Theorem 3.7). Following the proof of Tong et al. (2021a, Lemma 30), we can prove that $\mathbb{P}_r(\boldsymbol{X})$ is also the best tubal rank-r approximation of $\boldsymbol{X}$ under the partial Frobenius norm $\|\cdot\|_{F,r}$, as stated below.

Lemma 47 *Fix any $\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have*

$$\mathbb{P}_r(\boldsymbol{X}) = \underset{\widetilde{\boldsymbol{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}:\mathrm{rank}_t(\widetilde{\boldsymbol{X}}) \leq r}{\arg\min} \|\boldsymbol{X} - \widetilde{\boldsymbol{X}}\|_{F,r}.$$

Appendix B. Proof for Low-rank Tensor Factorization

B.1 Proof of Theorem 19

We prove Theorem 19 by induction. Specifically, we show that for all $t \geq 0$, (i) $\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \leq (1 - 0.7\eta)^t \mathrm{dist}(\boldsymbol{\mathcal{F}}_0, \boldsymbol{\mathcal{F}}_\star) \leq \frac{0.1}{\sqrt{\ell}}(1 - 0.7\eta)^t \bar{\sigma}_{s_r}(\boldsymbol{X}_\star)$ and (ii) the optimal alignment tensor $\boldsymbol{\mathcal{Q}}_t$ between $\boldsymbol{\mathcal{F}}_t$ and $\boldsymbol{\mathcal{F}}_\star$ exists.

For the base case $t = 0$, the first induction hypothesis on the distance metric trivially holds, and the second one also holds because of Lemma 41 and the assumption that $\mathrm{dist}(\boldsymbol{\mathcal{F}}_0, \boldsymbol{\mathcal{F}}_\star) \leq \frac{0.1}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\boldsymbol{X}_\star)$. Suppose that the t-th iterate $\boldsymbol{\mathcal{F}}_t$ obeys the aforementioned induction hypotheses. Our goal is to show that $\boldsymbol{\mathcal{F}}_{t+1}$ also satisfies these conditions. Let $\epsilon := 0.1$. By the definition of $\mathrm{dist}^2(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star)$, we have

$$\mathrm{dist}^2(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) \leq \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \|(\boldsymbol{\mathcal{R}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2.$$

According to the update rule (30), we have

$$(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}$$
$$= \left(\boldsymbol{\mathcal{L}}_\sharp - \eta(\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{X}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} - \boldsymbol{\mathcal{L}}_\star\right) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}$$
$$= \left(\boldsymbol{\mathcal{L}}_\triangle - \eta(\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H + \boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\triangle^H) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1}\right) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}$$
$$= (1 - \eta)\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} - \eta\boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\triangle^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}. \tag{49}$$

As a result, we can write (49) as $\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 = \frac{1}{\ell}\|(\bar{\boldsymbol{L}}_{t+1}\bar{\boldsymbol{Q}}_t - \bar{\boldsymbol{L}}_\star)\bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F^2$, where

$$\mathfrak{Q}_1 := \|(\bar{\boldsymbol{L}}_{t+1}\bar{\boldsymbol{Q}}_t - \bar{\boldsymbol{L}}_\star)\bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F^2$$
$$= (1 - \eta)^2 \mathrm{tr}(\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H) - 2\eta(1 - \eta)\mathrm{tr}(\bar{\boldsymbol{L}}_\star \bar{\boldsymbol{R}}_\triangle^H \bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H)$$
$$+ \eta^2 \mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{R}}_\triangle^H), \tag{50}$$

where we have used the fact $\bar{\boldsymbol{L}}_\star^H \bar{\boldsymbol{L}}_\star = \bar{\boldsymbol{G}}_\star$. Since $\boldsymbol{\mathcal{L}}_\sharp$ and $\boldsymbol{\mathcal{R}}_\sharp$ are aligned with $\boldsymbol{\mathcal{L}}_\star$ and $\boldsymbol{\mathcal{R}}_\star$, Lemma 42 tells that $\boldsymbol{\mathcal{G}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\triangle^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{L}}_\sharp = \boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star$, i.e., $\bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H \bar{\boldsymbol{L}}_\sharp = \bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star$. The

second term in (50) can be written as

$$
\begin{aligned}
&2\eta(1-\eta)\mathrm{tr}(\bar{\boldsymbol{L}}_\star \bar{\boldsymbol{R}}_\triangle^H \bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H) \\
=&2\eta(1-\eta)\mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H \bar{\boldsymbol{L}}_\star \bar{\boldsymbol{R}}_\triangle^H) \\
=&2\eta(1-\eta)\mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H \bar{\boldsymbol{L}}_\sharp \bar{\boldsymbol{R}}_\triangle^H) - 2\eta(1-\eta)\mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H \bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{R}}_\triangle^H) \\
=&2\eta(1-\eta)\mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{R}}_\triangle^H) - 2\eta(1-\eta)\mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H \bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{R}}_\triangle^H).
\end{aligned}
$$

The third term in (50) can be written as

$$
\begin{aligned}
&\eta^2\mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{G}}_\star (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{R}}_\triangle^H) \\
=&\eta^2\mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{R}}_\triangle^H) \\
&- \eta^2\mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}(\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp - \bar{\boldsymbol{G}}_\star)(\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{R}}_\triangle^H).
\end{aligned}
$$

Thus, (50) can be written in another form as

$$
\begin{aligned}
\mathfrak{Q}_1 =&(1-\eta)^2\mathrm{tr}(\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H) - \eta(2-3\eta)\mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{R}}_\triangle^H) \\
&+ 2\eta(1-\eta)\mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H \bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{R}}_\triangle^H) \\
&- \eta^2\mathrm{tr}(\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}(\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp - \bar{\boldsymbol{G}}_\star)(\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1}\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{R}}_\triangle^H).
\end{aligned}
$$

Following the proof in Tong et al. (2021a, Section B.2), we can bound this term by

$$
\begin{aligned}
\mathfrak{Q}_1 &\leq \left((1-\eta)^2 + \frac{2\epsilon}{1-\epsilon}\eta(1-\eta)\right)\mathrm{tr}(\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H) + \frac{2\epsilon+\epsilon^2}{(1-\epsilon)^2}\eta^2\mathrm{tr}(\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{R}}_\triangle^H) \\
&= \left((1-\eta)^2 + \frac{2\epsilon}{1-\epsilon}\eta(1-\eta)\right)\|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F^2 + \frac{2\epsilon+\epsilon^2}{(1-\epsilon)^2}\eta^2\|\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F^2.
\end{aligned}
$$

Hence,

$$
\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 \leq \left((1-\eta)^2 + \frac{2\epsilon}{1-\epsilon}\eta(1-\eta)\right)\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \frac{2\epsilon+\epsilon^2}{(1-\epsilon)^2}\eta^2\|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2.
$$

One can have a similar bound for the second term $\|(\boldsymbol{\mathcal{R}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2$. Therefore we obtain

$$
\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \|(\boldsymbol{\mathcal{R}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 \leq \hbar^2(\eta;\epsilon)\mathrm{dist}^2(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star),
$$

where the contraction rate is given by

$$
\hbar^2(\eta;\epsilon) := (1-\eta)^2 + \frac{2\epsilon}{1-\epsilon}\eta(1-\eta) + \frac{2\epsilon+\epsilon^2}{(1-\epsilon)^2}\eta^2.
$$

With $\epsilon = 0.1$ and $0 < \eta \leq \frac{2}{3}$, we have $\hbar(\eta;\epsilon) \leq 1 - 0.7\eta$. Thus we conclude that

$$
\begin{aligned}
\mathrm{dist}(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) &\leq \sqrt{\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \|(\boldsymbol{\mathcal{R}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2} \\
&\leq (1-0.7\eta)\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \\
&\leq (1-0.7\eta)^{t+1}\mathrm{dist}(\boldsymbol{\mathcal{F}}_0, \boldsymbol{\mathcal{F}}_\star) \leq (1-0.7\eta)^{t+1}\frac{0.1}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).
\end{aligned}
$$

This proves the first induction hypothesis. The existence of the optimal alignment tensor $\mathcal{Q}_{t+1}$ between $\mathcal{F}_{t+1}$ and $\mathcal{F}_\star$ is assured by Lemma 41, which finishes the proof for the second hypothesis.

The second conclusion is an easy consequence of Lemma 45 as

$$
\begin{aligned}
\|\mathcal{L}_t *_{\boldsymbol{\Phi}} \mathcal{R}_t^H - \mathcal{X}_\star\|_F &\leq (1 + \frac{\epsilon}{2})(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F + \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F) \\
&\leq (1 + \frac{\epsilon}{2})\sqrt{2}\mathrm{dist}(\mathcal{F}_t, \mathcal{F}_\star) \\
&\leq 1.5\mathrm{dist}(\mathcal{F}_t, \mathcal{F}_\star),
\end{aligned}
\tag{51}
$$

where we used $a + b \leq \sqrt{2(a^2 + b^2)}$ in the second line. The proof is now completed.

Appendix C. Proof for Tensor RPCA

Before presenting the proof of Lemma 20, we present several auxiliary lemmas.

Lemma 48 *If*

$$
\mathrm{dist}(\mathcal{F}_t, \mathcal{F}_\star) \leq \frac{\epsilon}{\sqrt{\ell}}\rho^t\bar{\sigma}_{s_r}(\mathcal{X}_\star)
$$

for some $0 < \epsilon < 1$, then the following inequalities hold

$$
\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \vee \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \leq \frac{\epsilon}{\sqrt{\ell}}\rho^t\bar{\sigma}_{s_r}(\mathcal{X}_\star);
$$

$$
\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \vee \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \leq \epsilon\rho^t\bar{\sigma}_{s_r}(\mathcal{X}_\star);
$$

$$
\|\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{L}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{L}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \vee \|\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \leq \frac{1}{1-\epsilon}.
$$

Proof Notice that $\mathrm{dist}(\mathcal{F}_t, \mathcal{F}_\star) \geq \sqrt{\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 \vee \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2}$. The first claim is directly followed by the definition of $\mathrm{dist}(\cdot, \cdot)$. By the fact that $\|\mathcal{A}\| = \|\bar{A}\| \leq \|\bar{A}\|_F = \|\bar{A}\|_F = \sqrt{\ell}\|\mathcal{A}\|_F$ for any tensor, the second claim can be deduced from the first claim. As a consequence, we have $\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\| \vee \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\| \leq \epsilon\rho^t \leq \epsilon$, given $\rho = 1 - 0.6\eta < 1$, then the third claim can be obtained by applying (41) and (42). ∎

Lemma 49 *If*

$$
\|(\mathcal{L}_{t+1} *_{\boldsymbol{\Phi}} \mathcal{Q}_t - \mathcal{L}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \vee \|(\mathcal{R}_{t+1} *_{\boldsymbol{\Phi}} \mathcal{Q}_t^{-H} - \mathcal{R}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \leq \epsilon\rho^{t+1}\bar{\sigma}_{s_r}(\mathcal{X}_\star),
$$

then

$$
\begin{aligned}
&\|\mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \mathcal{Q}_t^{-1} *_{\boldsymbol{\Phi}} (\mathcal{Q}_{t+1} - \mathcal{Q}_t) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \vee \|\mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \mathcal{Q}_t^H *_{\boldsymbol{\Phi}} (\mathcal{Q}_{t+1} - \mathcal{Q}_t)^{-H} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \\
&\leq \frac{2\epsilon}{1-\epsilon}\rho^{t+1}\bar{\sigma}_{s_r}(\mathcal{X}_\star).
\end{aligned}
$$

Proof First, following the proof of Tong et al. (2021a, Lemma 27), we have the following inequalities:

$$\|\mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \widetilde{\mathcal{Q}}^{-1} *_{\boldsymbol{\Phi}} \mathcal{Q} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} - \mathcal{G}_\star\| \leq \frac{\|\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\widetilde{\mathcal{Q}}^{-H} - \mathcal{Q}^{-H}) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|}{1 - \|(\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} \mathcal{Q}^{-H} - \mathcal{R}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\|},$$

$$\|\mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \widetilde{\mathcal{Q}}^{H} *_{\boldsymbol{\Phi}} \mathcal{Q}^{-H} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} - \mathcal{G}_\star\| \leq \frac{\|\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} (\widetilde{\mathcal{Q}} - \mathcal{Q}) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|}{1 - \|(\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{Q} - \mathcal{L}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\|}$$

for any $\mathcal{L}_\sharp \in \mathbb{R}^{n_1 \times r \times n_3}$, $\mathcal{R}_\sharp \in \mathbb{R}^{n_2 \times r \times n_3}$ and any invertible tensors $\mathcal{Q}, \widetilde{\mathcal{Q}} \in \mathrm{GL}(r)$, as long as $\|(\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{Q} - \mathcal{L}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\| \vee \|(\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} \mathcal{Q}^{-H} - \mathcal{R}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\| < 1$.

Next, notice that $\|(\mathcal{R}_{t+1} *_{\boldsymbol{\Phi}} \mathcal{Q}_{t+1}^{-H} - \mathcal{R}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \leq \epsilon \rho^{t+1} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$, i.e., $\|(\mathcal{R}_{t+1} *_{\boldsymbol{\Phi}} \mathcal{Q}_{t+1}^{-H} - \mathcal{R}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\| \leq \epsilon \rho^{t+1}$. Thus, by taking $\mathcal{R}_\sharp = \mathcal{R}_{t+1}$, $\mathcal{Q} = \mathcal{Q}_{t+1}$, and $\widetilde{\mathcal{Q}} = \mathcal{Q}_t$, we obtain

$$\|\mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \mathcal{Q}_t^{-1} *_{\boldsymbol{\Phi}} (\mathcal{Q}_{t+1} - \mathcal{Q}_t) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|$$

$$= \|\mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \mathcal{Q}_t^{-1} *_{\boldsymbol{\Phi}} \mathcal{Q}_{t+1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} - \mathcal{G}_\star\|$$

$$\leq \frac{\|\mathcal{R}_{t+1} *_{\boldsymbol{\Phi}} (\mathcal{Q}_t^{-H} - \mathcal{Q}_{t+1}^{-H}) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|}{1 - \|(\mathcal{R}_{t+1} *_{\boldsymbol{\Phi}} \mathcal{Q}_{t+1}^{-H} - \mathcal{R}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\|}$$

$$\leq \frac{\|(\mathcal{R}_{t+1} *_{\boldsymbol{\Phi}} \mathcal{Q}_t^{-H} - \mathcal{R}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| + \|(\mathcal{R}_{t+1} *_{\boldsymbol{\Phi}} \mathcal{Q}_{t+1}^{-H} - \mathcal{R}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|}{1 - \|(\mathcal{R}_{t+1} *_{\boldsymbol{\Phi}} \mathcal{Q}_{t+1}^{-H} - \mathcal{R}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\|}$$

$$\leq \frac{2\epsilon \rho^{t+1}}{1 - \epsilon \rho^{t+1}} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$$

$$\leq \frac{2\epsilon}{1 - \epsilon} \rho^{t+1} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star),$$

provided $\rho = 1 - 0.6\eta < 1$. Similarly, one can see

$$\|\mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \mathcal{Q}_t^{H} *_{\boldsymbol{\Phi}} (\mathcal{Q}_{t+1} - \mathcal{Q}_t)^{-H} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \leq \frac{2\epsilon}{1 - \epsilon} \rho^{t+1} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).$$

This finishes the proof. ∎

Lemma 50 *If*

$$\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \leq \frac{\epsilon}{\sqrt{\ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) \quad \text{and}$$

$$\sqrt{n_1} \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2} \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star),$$

then

$$\|\boldsymbol{\mathcal{X}}_\star - \boldsymbol{\mathcal{X}}_t\|_\infty \leq 3 \frac{\mu s_r}{n_3 \sqrt{n_1 n_2} \ell} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).$$

Proof First, based on Definition 10 and the assumption of this lemma, we have

$$\|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \le \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}\|\mathcal{G}_\star^{-1}\| + \|\mathcal{R}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty}$$
$$\le (\rho^t + 1)\sqrt{\frac{\mu s_r}{n_2 n_3 \ell}} \le 2\sqrt{\frac{\mu s_r}{n_2 n_3 \ell}}.$$

Furthermore, using Lemma 35, we obtain

$$\|\mathcal{X}_\star - \mathcal{X}_t\|_\infty = \|\mathcal{L}_\triangle *_\Phi \mathcal{R}_\sharp^H + \mathcal{L}_\star *_\Phi \mathcal{R}_\triangle^H\|_\infty \le \|\mathcal{L}_\triangle *_\Phi \mathcal{R}_\sharp^H\|_\infty + \|\mathcal{L}_\star *_\Phi \mathcal{R}_\triangle^H\|_\infty$$
$$\le \sqrt{\ell}\|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}\|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} + \sqrt{\ell}\|\mathcal{L}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty}\|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$
$$\le \sqrt{\ell}(2\sqrt{\frac{\mu s_r}{n_2 n_3 \ell}} + \sqrt{\frac{\mu s_r}{n_2 n_3 \ell}})\sqrt{\frac{\mu s_r}{n_1 n_3 \ell}}\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star)$$
$$= 3\frac{\mu s_r}{n_3\sqrt{n_1 n_2}\ell}\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star).$$

This finishes the proof. ∎

C.1 Proof of Lemma 20

This proof is done by induction.

Base case. Since $\rho^0 = 1$, the assumed initial conditions satisfy the base case at $t = 0$.

Induction step. At the t-th iteration, we assume the conditions

$$\text{dist}(\mathcal{F}_t, \mathcal{F}_\star) \le \frac{\epsilon}{\sqrt{\ell}}\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star),$$

$$\sqrt{n_1}\|(\mathcal{L}_t *_\Phi \mathcal{Q}_t - \mathcal{L}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|(\mathcal{R}_t *_\Phi \mathcal{Q}_t^{-H} - \mathcal{R}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \le \sqrt{\frac{\mu s_r}{n_3 \ell}}\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star)$$

hold. In view of the condition $\text{dist}(\mathcal{F}_t, \mathcal{F}_\star) \le \frac{0.02}{\sqrt{\ell}}\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star)$ and Lemma 41, one knows that $\mathcal{Q}_t$, the optimal alignment tensor between $\mathcal{F}_t$ and $\mathcal{F}_\star$ exists, and $\epsilon := 0.02$. In what follows, we shall prove the distance contraction and the incoherence condition separately.

Distance contraction: By the definition of $\text{dist}^2(\mathcal{F}_{t+1}, \mathcal{F}_\star)$, we have

$$\text{dist}^2(\mathcal{F}_{t+1}, \mathcal{F}_\star) \le \|(\mathcal{L}_{t+1} *_\Phi \mathcal{Q}_t - \mathcal{L}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \|(\mathcal{R}_{t+1} *_\Phi \mathcal{Q}_t^{-H} - \mathcal{R}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F^2.$$

According to the update rule (15), we have

$$
\begin{aligned}
&(\mathcal{L}_{t+1} *_{\Phi} \mathcal{Q}_t - \mathcal{L}_\star) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}} \\
=&\Big(\mathcal{L}_\sharp - \eta(\mathcal{L}_\sharp *_{\Phi} \mathcal{R}_\sharp^H + \mathcal{S}_{t+1} - \mathcal{X}_\star - \mathcal{S}_\star) *_{\Phi} \mathcal{R}_\sharp *_{\Phi} (\mathcal{R}_\sharp^H *_{\Phi} \mathcal{R}_\sharp)^{-1} - \mathcal{L}_\star\Big) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}} \\
=&\Big(\mathcal{L}_\triangle - \eta(\mathcal{L}_\sharp *_{\Phi} \mathcal{R}_\sharp^H - \mathcal{X}_\star) *_{\Phi} \mathcal{R}_\sharp *_{\Phi} (\mathcal{R}_\sharp^H *_{\Phi} \mathcal{R}_\sharp)^{-1} \\
&\qquad\qquad\qquad - \eta(\mathcal{S}_{t+1} - \mathcal{S}_\star) *_{\Phi} \mathcal{R}_\sharp *_{\Phi} (\mathcal{R}_\sharp^H *_{\Phi} \mathcal{R}_\sharp)^{-1}\Big) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}} \\
=&\Big(\mathcal{L}_\triangle - \eta(\mathcal{L}_\triangle *_{\Phi} \mathcal{R}_\sharp^H + \mathcal{L}_\star *_{\Phi} \mathcal{R}_\triangle^H) *_{\Phi} \mathcal{R}_\sharp *_{\Phi} (\mathcal{R}_\sharp^H *_{\Phi} \mathcal{R}_\sharp)^{-1} \\
&\qquad\qquad\qquad - \eta\mathcal{S}_\triangle *_{\Phi} \mathcal{R}_\sharp *_{\Phi} (\mathcal{R}_\sharp^H *_{\Phi} \mathcal{R}_\sharp)^{-1}\Big) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}} \\
=&(1-\eta)\mathcal{L}_\triangle *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}} - \eta\mathcal{L}_\star *_{\Phi} \mathcal{R}_\triangle^H *_{\Phi} \mathcal{R}_\sharp *_{\Phi} (\mathcal{R}_\sharp^H *_{\Phi} \mathcal{R}_\sharp)^{-1} *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}} \\
&\qquad\qquad\qquad - \eta\mathcal{S}_\triangle *_{\Phi} \mathcal{R}_\sharp *_{\Phi} (\mathcal{R}_\sharp^H *_{\Phi} \mathcal{R}_\sharp)^{-1} *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}.
\end{aligned}
\tag{52}
$$

Taking the squared Frobenius norm of both sides of (52) to obtain

$$
\begin{aligned}
\|(\mathcal{L}_{t+1} *_{\Phi} \mathcal{Q}_t - \mathcal{L}_\star) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 =&\frac{1}{\ell}\|(\bar{L}_{t+1}\bar{Q}_t - \bar{L}_\star)\bar{G}_\star^{\frac{1}{2}}\|_F^2 \\
=&\frac{1}{\ell}\Big(\|(1-\eta)\bar{L}_\triangle\bar{G}_\star^{\frac{1}{2}} - \eta\bar{L}_\star\bar{R}_\triangle^H\bar{R}_\sharp(\bar{R}_\sharp^H\bar{R}_\sharp)^{-1}\bar{G}_\star^{\frac{1}{2}}\|_F^2 \\
&\quad - 2\eta(1-\eta)\mathrm{tr}\Big(\bar{S}_\triangle\bar{R}_\sharp(\bar{R}_\sharp^H\bar{R}_\sharp)^{-1}\bar{G}_\star\bar{L}_\triangle^H\Big) \\
&\quad + 2\eta^2\mathrm{tr}\Big(\bar{S}_\triangle\bar{R}_\sharp(\bar{R}_\sharp^H\bar{R}_\sharp)^{-1}\bar{G}_\star(\bar{R}_\sharp^H\bar{R}_\sharp)^{-1}\bar{R}_\sharp^H\bar{R}_\triangle\bar{L}_\star^H\Big) \\
&\quad + \eta^2\|\bar{S}_\triangle\bar{R}_\sharp(\bar{R}_\sharp^H\bar{R}_\sharp)^{-1}\bar{G}_\star^{\frac{1}{2}}\|_F^2\Big) \\
:=&\frac{1}{\ell}(\mathfrak{R}_1 - \mathfrak{R}_2 + \mathfrak{R}_3 + \mathfrak{R}_4).
\end{aligned}
$$

Bound of $\mathfrak{R}_1$. The component $\mathfrak{R}_1$ is identical to (50), and the bound of this term was shown therein. We will clear this bound further by applying Lemma 48, that is

$$
\begin{aligned}
\mathfrak{R}_1 \leq& \Big((1-\eta)^2 + \frac{2\epsilon}{1-\epsilon}\eta(1-\eta)\Big)\mathrm{tr}(\bar{L}_\triangle\bar{G}_\star\bar{L}_\triangle^H) + \frac{2\epsilon + \epsilon^2}{(1-\epsilon)^2}\eta^2\mathrm{tr}(\bar{R}_\triangle\bar{G}_\star\bar{R}_\triangle^H) \\
=& (1-\eta)^2\|\bar{L}_\triangle\bar{G}_\star^{\frac{1}{2}}\|_F^2 + \frac{2\epsilon}{1-\epsilon}\eta(1-\eta)\|\bar{L}_\triangle\bar{G}_\star^{\frac{1}{2}}\|_F^2 + \frac{2\epsilon + \epsilon^2}{(1-\epsilon)^2}\eta^2\|\bar{R}_\triangle\bar{G}_\star^{\frac{1}{2}}\|_F^2 \\
=& (1-\eta)^2\|\bar{L}_\triangle\bar{G}_\star^{\frac{1}{2}}\|_F^2 + \Big((1-\eta)\frac{2\epsilon^3}{1-\epsilon} + \eta\frac{2\epsilon^3 + \epsilon^4}{(1-\epsilon)^2}\Big)\eta\rho^{2t}\bar{\sigma}_{sr}^2(\mathcal{X}_\star).
\end{aligned}
\tag{53}
$$

Bound of $\mathfrak{R}_2$. According to Lemma 34, $\|\bar{S}_\triangle\| = \|\mathcal{S}_\triangle\| \leq \frac{\alpha\sqrt{\ell}}{2}(n_1 + n_2 n_3)\|\mathcal{S}_\triangle\|_\infty$. Lemma 22 implies $\mathcal{S}_\triangle$ is an α-sparse tensor and our threshold ζ_{t+1} suggests that $\|\mathcal{S}_\triangle\|_\infty \leq$

$6\frac{\mu s_r}{n_3\sqrt{n_1 n_2 \ell}}\rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$, we further have

$$
\begin{aligned}
&\left|\mathrm{tr}\left(\bar{\boldsymbol{S}}_\triangle \bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H\right)\right|\\
&\leq \|\bar{\boldsymbol{S}}_\triangle\| \|\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H\|_*\\
&\leq \frac{\alpha\sqrt{\ell}}{2}(n_1 + n_2 n_3)\|\boldsymbol{\mathcal{S}}_\triangle\|_\infty \sqrt{s_r}\|\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{L}}_\triangle^H\|_F\\
&\leq \frac{\alpha\sqrt{\ell}}{2}(n_1 + n_2 n_3)\|\boldsymbol{\mathcal{S}}_\triangle\|_\infty \sqrt{s_r}\|\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\| \|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F\\
&\leq 3\alpha\mu s_r^{1.5}\frac{n_1 + n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{\epsilon}{1-\epsilon}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star).
\end{aligned}
$$

Hence,

$$
|\mathfrak{R}_2| \leq 6\eta(1-\eta)\alpha\mu s_r^{1.5}\frac{n_1 + n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{\epsilon}{1-\epsilon}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star).
$$

Bound of $\mathfrak{R}_3$. Similar to $\mathfrak{R}_2$, we have

$$
\begin{aligned}
&\left|\mathrm{tr}\left(\bar{\boldsymbol{S}}_\triangle \bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{L}}_\star^H\right)\right|\\
&\leq \|\bar{\boldsymbol{S}}_\triangle\| \|\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{L}}_\star^H\|_*\\
&\leq \frac{\alpha\sqrt{\ell}}{2}(n_1 + n_2 n_3)\|\boldsymbol{\mathcal{S}}_\triangle\|_\infty \sqrt{s_r}\|\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{L}}_\star^H\|_F\\
&\leq \frac{\alpha\sqrt{\ell}}{2}(n_1 + n_2 n_3)\|\boldsymbol{\mathcal{S}}_\triangle\|_\infty \sqrt{s_r}\|\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|^2 \|\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{L}}_\star^H\|_F\\
&\leq \frac{\alpha\sqrt{\ell}}{2}(n_1 + n_2 n_3)\|\boldsymbol{\mathcal{S}}_\triangle\|_\infty \sqrt{s_r}\|\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|^2 \|\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F \|\bar{\boldsymbol{U}}_\star\|\\
&\leq 3\alpha\mu s_r^{1.5}\frac{n_1 + n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{\epsilon}{(1-\epsilon)^2}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star).
\end{aligned}
$$

Hence,

$$
|\mathfrak{R}_3| \leq 6\eta^2\alpha\mu s_r^{1.5}\frac{n_1 + n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{\epsilon}{(1-\epsilon)^2}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star).
$$

Bound of $\mathfrak{R}_4$.

$$
\begin{aligned}
\|\bar{\boldsymbol{S}}_\triangle \bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F^2 &\leq s_r\|\bar{\boldsymbol{S}}_\triangle \bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|^2\\
&\leq s_r\|\bar{\boldsymbol{S}}_\triangle\|^2 \|\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|^2\\
&\leq \frac{\alpha^2 \ell s_r}{4}(n_1 + n_2 n_3)^2\|\boldsymbol{\mathcal{S}}_\triangle\|_\infty^2 \|\bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|^2\\
&\leq 9\alpha^2\mu^2 s_r^3 \frac{(n_1 + n_2 n_3)^2}{n_1 n_2 n_3^2}\frac{1}{(1-\epsilon)^2}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star).
\end{aligned}
$$

Hence,

$$
|\mathfrak{R}_4| \leq 9\eta^2\alpha^2\mu^2 s_r^3 \frac{(n_1 + n_2 n_3)^2}{n_1 n_2 n_3^2}\frac{1}{(1-\epsilon)^2}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star).
$$

Combining all the bounds together, we have

$$
\begin{aligned}
&\|(\mathcal{L}_{t+1} *_{\mathbf{\Phi}} \mathcal{Q}_t - \mathcal{L}_\star) *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 \\
&\leq \frac{1}{\ell}\Big((1-\eta)^2\|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F^2 + \Big((1-\eta)\frac{2\epsilon^3}{1-\epsilon} + \eta\frac{2\epsilon^3+\epsilon^4}{(1-\epsilon)^2}\Big)\eta\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star) \\
&\quad + 6\eta(1-\eta)\alpha\mu s_r^{1.5}\frac{n_1+n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{\epsilon}{1-\epsilon}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star) \\
&\quad + 6\eta^2\alpha\mu s_r^{1.5}\frac{n_1+n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{\epsilon}{(1-\epsilon)^2}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star) \\
&\quad + 9\eta^2\alpha^2\mu^2 s_r^3\frac{(n_1+n_2 n_3)^2}{n_1 n_2 n_3^2}\frac{1}{(1-\epsilon)^2}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star)\Big).
\end{aligned}
$$

We can obtain a similar bound for $\|(\mathcal{R}_{t+1} *_{\mathbf{\Phi}} \mathcal{Q}_t^{-H} - \mathcal{R}_\star) *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2$. Thus, we have

$$
\begin{aligned}
&\mathrm{dist}^2(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) \\
&\leq \frac{1}{\ell}\Big((1-\eta)^2\Big(\|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F^2 + \|\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F^2\Big) + 2\Big((1-\eta)\frac{2\epsilon^3}{1-\epsilon} + \eta\frac{2\epsilon^3+\epsilon^4}{(1-\epsilon)^2}\Big)\eta\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star) \\
&\quad + 12\eta(1-\eta)\alpha\mu s_r^{1.5}\frac{n_1+n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{\epsilon}{1-\epsilon}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star) + 12\eta^2\alpha\mu s_r^{1.5}\frac{n_1+n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{\epsilon}{(1-\epsilon)^2}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star) \\
&\quad + 18\eta^2\alpha^2\mu^2 s_r^3\frac{(n_1+n_2 n_3)^2}{n_1 n_2 n_3^2}\frac{1}{(1-\epsilon)^2}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star)\Big) \\
&\leq \Big((1-\eta)^2 + 2\Big((1-\eta)\frac{2\epsilon}{1-\epsilon} + \eta\frac{2\epsilon+\epsilon^2}{(1-\epsilon)^2}\Big)\eta + 12\eta(1-\eta)\alpha\mu s_r^{1.5}\frac{n_1+n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{1}{\epsilon(1-\epsilon)} \\
&\quad + 12\eta^2\alpha\mu s_r^{1.5}\frac{n_1+n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{1}{\epsilon(1-\epsilon)^2} + 18\eta^2\alpha^2\mu^2 s_r^3\frac{(n_1+n_2 n_3)^2}{n_1 n_2 n_3^2}\frac{1}{\epsilon^2(1-\epsilon)^2}\Big)\frac{1}{\ell}\epsilon^2\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star) \\
&\leq (1-0.6\eta)^2\frac{1}{\ell}\epsilon^2\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star), \tag{54}
\end{aligned}
$$

where we use the fact $\|\bar{\boldsymbol{L}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F^2 + \|\bar{\boldsymbol{R}}_\triangle \bar{\boldsymbol{G}}_\star^{\frac{1}{2}}\|_F^2 = \ell\,\mathrm{dist}^2(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \leq \epsilon^2\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star)$ in the second step, and the last step holds with $\epsilon = 0.02$, $\alpha\mu s_r^{1.5}\frac{n_1+n_2 n_3}{n_3\sqrt{n_1 n_2}} \leq \alpha\mu s_r^{1.5}\frac{n_1+n_2 n_3}{n_3\sqrt{n_{(2)} n_2}} \leq 10^{-4}$, and $\frac{1}{4} < \eta \leq \frac{8}{9}$. Thus we conclude that

$$
\mathrm{dist}(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) \leq \frac{\epsilon}{\sqrt{\ell}}\rho^{t+1}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)
$$

by setting $\rho = 1 - 0.6\eta$.

Incoherence condition: We first use (52) again to obtain

$$
\begin{aligned}
&\|(\mathcal{L}_{t+1} *_{\mathbf{\Phi}} \mathcal{Q}_t - \mathcal{L}_\star) *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \\
&\leq (1-\eta)\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} + \eta\|\mathcal{L}_\star *_{\mathbf{\Phi}} \mathcal{R}_\triangle^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \\
&\quad + \eta\|\mathcal{S}_\triangle *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \\
&:= \mathfrak{B}_1 + \mathfrak{B}_2 + \mathfrak{B}_3.
\end{aligned}
$$

Bound of $\mathfrak{B}_1$. Followed by the assumption in this step, we can easily have $\mathfrak{B}_1 \leq (1 - \eta)\sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$.

Bound of $\mathfrak{B}_2$. Definition 10 implies $\|\boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}}$. Thus, using Lemma 36 and Lemma 48, we have

$$
\begin{aligned}
\mathfrak{B}_2 &\leq \eta \|\boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\| \|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\| \\
&\leq \eta \frac{\epsilon}{1-\epsilon} \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).
\end{aligned}
$$

Bound of $\mathfrak{B}_3$. By Lemma 22, $\mathrm{supp}(\boldsymbol{\mathcal{S}}_\triangle) \subseteq \mathrm{supp}(\boldsymbol{\mathcal{S}}_\star)$, which implies that $\boldsymbol{\mathcal{S}}_\triangle$ is an α-sparse tensor. Thus, using Lemmas 34, 22 and 50, we have

$$
\begin{aligned}
\mathfrak{B}_3 &\leq \eta \|\boldsymbol{\mathcal{S}}_\triangle\|_{2,\infty} \|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\| \leq \eta \frac{\sqrt{\alpha n_2 n_3}}{1-\epsilon} \|\boldsymbol{\mathcal{S}}_\triangle\|_\infty \\
&\leq 6\eta \frac{\sqrt{\alpha \mu s_r}}{1-\epsilon} \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).
\end{aligned}
$$

Putting all the bounds together, we obtain

$$
\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \left((1-\eta) + \eta \frac{\epsilon}{1-\epsilon} + 6\eta \frac{\sqrt{\alpha \mu s_r}}{1-\epsilon} \right) \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).
$$

We can then have

$$
\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|_{2,\infty} \leq \left((1-\eta) + \eta \frac{\epsilon}{1-\epsilon} + 6\eta \frac{\sqrt{\alpha \mu s_r}}{1-\epsilon} \right) \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t.
$$

The last step is to switch the alignment tensor from $\boldsymbol{\mathcal{Q}}_t$ to $\boldsymbol{\mathcal{Q}}_{t+1}$. Note that (54) together with Lemma 41 confirms the existence of $\boldsymbol{\mathcal{Q}}_{t+1}$. Applying the triangle inequality and

Lemma 49, we have

$$\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_{t+1} - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty}$$

$$\leq \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty} + \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{Q}}_{t+1} - \boldsymbol{\mathcal{Q}}_t) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty}$$

$$= \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty}$$

$$+ \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-1} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{Q}}_{t+1} - \boldsymbol{\mathcal{Q}}_t) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty}$$

$$\leq \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty}$$

$$+ \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-1} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{Q}}_{t+1} - \boldsymbol{\mathcal{Q}}_t) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|$$

$$\leq \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty} + \Big(\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_{2,\infty}$$

$$+ \|\boldsymbol{\mathcal{L}}_{\star} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_{2,\infty} \Big) \|\boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-1} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{Q}}_{t+1} - \boldsymbol{\mathcal{Q}}_t) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|$$

$$\leq \Big((1-\eta) + \eta \frac{\epsilon}{1-\epsilon} + 6\eta \frac{\sqrt{\alpha \mu s_r}}{1-\epsilon} \Big) \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_{\star})$$

$$+ \Big(((1-\eta) + \eta \frac{\epsilon}{1-\epsilon} + 6\eta \frac{\sqrt{\alpha \mu s_r}}{1-\epsilon}) \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t + \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \Big) \frac{2\epsilon}{1-\epsilon} \rho^{t+1} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_{\star})$$

$$\leq \Big((1-\eta) + \eta \frac{\epsilon}{1-\epsilon} + 6\eta \frac{\sqrt{\alpha \mu s_r}}{1-\epsilon} + \frac{2\epsilon}{1-\epsilon}((2-\eta) + \eta \frac{\epsilon}{1-\epsilon} + 6\eta \frac{\sqrt{\alpha \mu s_r}}{1-\epsilon}) \Big) \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_{\star})$$

$$\leq (1 - 0.6\eta) \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_{\star}),$$

where we use $\epsilon = 0.02$, $\alpha \mu s_r < \alpha \mu s_r^{1.5} \frac{n_1 + n_2 n_3}{n_3 \sqrt{n_{(2)} n_2}} \leq 10^{-4}$, and $\frac{1}{4} \leq \eta \leq \frac{8}{9}$ in the last step. A similar result can be computed for $\|(\boldsymbol{\mathcal{R}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_{t+1}^{-H} - \boldsymbol{\mathcal{R}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty}$. The conclusion that

$$\sqrt{n_1} \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_{t+1} - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2} \|(\boldsymbol{\mathcal{R}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_{t+1}^{-H} - \boldsymbol{\mathcal{R}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty}$$

$$\leq \sqrt{\frac{\mu s_r}{n_3 \ell}} \rho^{t+1} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_{\star})$$

can be achieved by setting $\rho = 1 - 0.6\eta$. This finishes the proof.

C.2 Proof of Lemma 21

First, notice that the (i,j)-th mode-3 tube of $\bar{\mathbf{e}}_{ijk}$, which is equal to $L(\dot{\mathbf{e}}_k)$, is the only nonzero mode-3 tube of $\bar{\mathbf{e}}_{ijk}$. Thus the only nonzero mode-3 tube of $L(\bar{\mathbf{e}}_{ijk})$ is its (i,j)-th mode-3 tube, and it is equal to $L(L(\dot{\mathbf{e}}_k))$. Then $L(\bar{\mathbf{e}}_{ijk})$ can be written as $L(\bar{\mathbf{e}}_{ijk}) =$

$L(\mathring{\mathbf{e}}_i) \triangle L(L(\dot{\mathbf{e}}_k)) \triangle L(\mathring{\mathbf{e}}_j^H)$. By Definition 10, we have

$$
\begin{aligned}
\|\boldsymbol{\mathcal{X}}_\star\|_\infty &= \max_{i,j,k} |\langle \boldsymbol{\mathcal{U}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}_\star^H, \bar{\mathbf{e}}_{ijk}\rangle| = \frac{1}{\ell} \max_{i,j,k} |\langle \bar{\boldsymbol{\mathcal{U}}}_\star \triangle \bar{\boldsymbol{\mathcal{G}}}_\star \triangle \bar{\boldsymbol{\mathcal{V}}}_\star^H, L(\bar{\mathbf{e}}_{ijk})\rangle| \\
&= \frac{1}{\ell} \max_{i,j,k} |\langle \bar{\boldsymbol{\mathcal{U}}}_\star \triangle \bar{\boldsymbol{\mathcal{G}}}_\star \triangle \bar{\boldsymbol{\mathcal{V}}}_\star^H, L(\mathring{\mathbf{e}}_i) \triangle L(L(\dot{\mathbf{e}}_k)) \triangle L(\mathring{\mathbf{e}}_j^H)\rangle| \\
&= \frac{1}{\ell} \max_{i,j,k} |\langle \bar{\boldsymbol{U}}_\star \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{V}}_\star^H, \bar{\mathring{\boldsymbol{e}}}_i \bar{\boldsymbol{h}}_k \bar{\mathring{\boldsymbol{e}}}_j^H \rangle| = \frac{1}{\ell} \max_{i,j,k} |\langle \bar{\mathring{\boldsymbol{e}}}_i^H \bar{\boldsymbol{U}}_\star \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{V}}_\star^H \bar{\mathring{\boldsymbol{e}}}_j, \bar{\boldsymbol{h}}_k\rangle| \\
&\leq \frac{1}{\ell} \max_{i,j,k} \|\bar{\mathring{\boldsymbol{e}}}_i^H \bar{\boldsymbol{U}}_\star \bar{\boldsymbol{G}}_\star \bar{\boldsymbol{V}}_\star^H \bar{\mathring{\boldsymbol{e}}}_j\|_F \|\bar{\boldsymbol{h}}_k\|_F \\
&\leq \frac{1}{\ell} \max_{i,j,k} \|\bar{\mathring{\boldsymbol{e}}}_i^H \bar{\boldsymbol{U}}_\star\|_F \|\bar{\boldsymbol{G}}_\star\| \|\bar{\boldsymbol{V}}_\star^H \bar{\mathring{\boldsymbol{e}}}_j\|_F \|\bar{\boldsymbol{h}}_k\|_F \\
&\leq \max_{i,j,k} \|\mathring{\mathbf{e}}_i^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{U}}_\star\|_F \|\bar{\boldsymbol{G}}_\star\| \|\boldsymbol{\mathcal{V}}_\star^H *_{\boldsymbol{\Phi}} \mathring{\mathbf{e}}_j\|_F \|\bar{\boldsymbol{h}}_k\|_F \\
&\leq \frac{\mu s_r}{n_3 \sqrt{n_1 n_2 \ell}} \bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star),
\end{aligned}
$$

where $\bar{\mathring{\boldsymbol{e}}}_i = \mathtt{bdiag}(L(\mathring{\mathbf{e}}_i))$ and $\mathfrak{h}_k = L(L(\dot{\mathbf{e}}_k))$. Invoking Lemma 22 with $\boldsymbol{\mathcal{X}}_{-1} = \boldsymbol{0}$, we have

$$
\|\boldsymbol{\mathcal{S}}_\star - \boldsymbol{\mathcal{S}}_0\|_\infty \leq 2 \frac{\mu s_r}{n_3 \sqrt{n_1 n_2 \ell}} \bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star) \quad \text{and} \quad \mathrm{supp}(\boldsymbol{\mathcal{S}}_0) \subseteq \mathrm{supp}(\boldsymbol{\mathcal{S}}_\star),
$$

which implies $\boldsymbol{\mathcal{S}}_\star - \boldsymbol{\mathcal{S}}_0$ is an α-sparse tensor. Applying Lemma 34, we have

$$
\|\boldsymbol{\mathcal{S}}_\star - \boldsymbol{\mathcal{S}}_0\| \leq \frac{\alpha\sqrt{\ell}}{2}(n_1 + n_2 n_3)\|\boldsymbol{\mathcal{S}}_\star - \boldsymbol{\mathcal{S}}_0\|_\infty \leq \alpha \mu s_r \frac{n_1 + n_2 n_3}{n_3 \sqrt{n_1 n_2}} \bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star) = \alpha \mu s_r \kappa \frac{n_1 + n_2 n_3}{n_3 \sqrt{n_1 n_2}} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).
$$

Since $\boldsymbol{\mathcal{X}}_0 = \boldsymbol{\mathcal{L}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_0^H$ is the best approximation of $\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{S}}_0$ with tubal rank r, we obtain

$$
\begin{aligned}
\|\boldsymbol{\mathcal{X}}_\star - \boldsymbol{\mathcal{X}}_0\| = \|\bar{\boldsymbol{X}}_\star - \bar{\boldsymbol{X}}_0\| &\leq \|\bar{\boldsymbol{X}}_\star - (\bar{\boldsymbol{Y}} - \bar{\boldsymbol{S}}_0)\| + \|(\bar{\boldsymbol{Y}} - \bar{\boldsymbol{S}}_0) - \bar{\boldsymbol{X}}_0\| \\
&\leq 2\|\bar{\boldsymbol{X}}_\star - (\bar{\boldsymbol{Y}} - \bar{\boldsymbol{S}}_0)\| \\
&= 2\|\bar{\boldsymbol{S}}_\star - \bar{\boldsymbol{S}}_0\| \\
&\leq 2\alpha \mu s_r \kappa \frac{n_1 + n_2 n_3}{n_3 \sqrt{n_1 n_2}} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star),
\end{aligned}
$$

where we use the definition $\bar{\boldsymbol{Y}} = \bar{\boldsymbol{X}}_\star + \bar{\boldsymbol{S}}_\star$ in the equality. Using Lemma 43, we obtain

$$
\begin{aligned}
\mathrm{dist}(\boldsymbol{\mathcal{F}}_0, \boldsymbol{\mathcal{F}}_\star) &\leq (\sqrt{2}+1)^{\frac{1}{2}} \frac{1}{\sqrt{\ell}} \|\bar{\boldsymbol{X}}_\star - \bar{\boldsymbol{X}}_0\|_F \leq \sqrt{\frac{2(\sqrt{2}+1)s_r}{\ell}} \|\bar{\boldsymbol{X}}_\star - \bar{\boldsymbol{X}}_0\| \\
&\leq 5\frac{1}{\sqrt{\ell}} \alpha \mu s_r^{1.5} \kappa \frac{n_1 + n_2 n_3}{n_3 \sqrt{n_1 n_2}} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star),
\end{aligned}
$$

where we use the fact that $\bar{\boldsymbol{X}}_\star - \bar{\boldsymbol{X}}_0$ has at most rank-$2s_r$. Given $\epsilon = 5c_0$ and $\alpha \leq \frac{c_0}{\mu s_r^{1.5} \kappa \frac{n_1 + n_2 n_3}{n_3 \sqrt{n_1 n_2}}}$, our first claim

$$
\mathrm{dist}(\boldsymbol{\mathcal{F}}_0, \boldsymbol{\mathcal{F}}_\star) \leq \frac{5c_0}{\sqrt{\ell}} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) \tag{55}
$$

is proved.

Next, we proceed to prove the second claim. For the ease of presentation, we define $\mathcal{L}_\sharp := \mathcal{L}_0 *_\Phi \mathcal{Q}_0$, $\mathcal{R}_\sharp := \mathcal{R}_0 *_\Phi \mathcal{Q}_0^{-H}$, $\mathcal{L}_\triangle := \mathcal{L}_\sharp - \mathcal{L}_\star$, $\mathcal{R}_\triangle := \mathcal{R}_\sharp - \mathcal{R}_\star$, $\mathcal{S}_\triangle := \mathcal{S}_0 - \mathcal{S}_\star$. We first notice that $\mathcal{U}_0 *_\Phi \mathcal{G}_0 *_\Phi \mathcal{V}_0^H = \text{t-SVD}_r(\mathcal{Y} - \mathcal{S}_0) = \text{t-SVD}_r(\mathcal{X}_\star - \mathcal{S}_\triangle)$, thus

$$\mathcal{L}_0 = \mathcal{U}_0 *_\Phi \mathcal{G}_0^{\frac{1}{2}} = (\mathcal{X}_\star - \mathcal{S}_\triangle) *_\Phi \mathcal{V}_0 *_\Phi \mathcal{G}_0^{-\frac{1}{2}} = (\mathcal{X}_\star - \mathcal{S}_\triangle) *_\Phi \mathcal{R}_0 *_\Phi \mathcal{G}_0^{-1}$$
$$= (\mathcal{X}_\star - \mathcal{S}_\triangle) *_\Phi \mathcal{R}_0 *_\Phi (\mathcal{R}_0^H *_\Phi \mathcal{R}_0)^{-1}.$$

Multiplying $\mathcal{Q}_0 *_\Phi \mathcal{G}_\star^{\frac{1}{2}}$ on both sides using transformed t-product, we have

$$\mathcal{L}_\sharp *_\Phi \mathcal{G}_\star^{\frac{1}{2}} = \mathcal{L}_0 *_\Phi \mathcal{Q}_0 *_\Phi \mathcal{G}_\star^{\frac{1}{2}} = (\mathcal{X}_\star - \mathcal{S}_\triangle) *_\Phi \mathcal{R}_0 *_\Phi (\mathcal{R}_0^H *_\Phi \mathcal{R}_0)^{-1} *_\Phi \mathcal{Q}_0 *_\Phi \mathcal{G}_\star^{\frac{1}{2}}$$
$$= (\mathcal{X}_\star - \mathcal{S}_\triangle) *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}.$$

Subtracting $\mathcal{X}_\star *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}$ on both sides and using the fact that $\mathcal{L}_\sharp *_\Phi \mathcal{G}_\star^{\frac{1}{2}} = \mathcal{L}_\sharp *_\Phi \mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}$, along with the decomposition that $\mathcal{L}_\sharp *_\Phi \mathcal{R}_\sharp^H - \mathcal{X}_\star = \mathcal{L}_\triangle *_\Phi \mathcal{R}_\sharp^H + \mathcal{L}_\star *_\Phi \mathcal{R}_\triangle^H$, we have

$$\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}} + \mathcal{L}_\star *_\Phi \mathcal{R}_\triangle^H *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}} = -\mathcal{S}_\triangle *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}.$$

Thus,

$$\|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \|\mathcal{L}_\star *_\Phi \mathcal{R}_\triangle^H *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$
$$+ \|\mathcal{S}_\triangle *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$
$$:= \mathfrak{J}_1 + \mathfrak{J}_2.$$

Bound of $\mathfrak{J}_1$. Since (55) holds, Lemma 48 implies $\|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \leq \epsilon \bar{\sigma}_{s_r}(\mathcal{X}_\star)$ and $\|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \leq \frac{1}{1-\epsilon}$. By Definition 10 and Lemma 36, we obtain

$$\mathfrak{J}_1 \leq \|\mathcal{L}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|$$
$$\leq \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \frac{\epsilon}{1-\epsilon} \bar{\sigma}_{s_r}(\mathcal{X}_\star). \tag{56}$$

Bound of $\mathfrak{J}_2$. By Lemmas 34, 36 and 37, we have

$$\mathfrak{J}_2 \leq \sqrt{n_2 \ell} \|\mathcal{S}_\triangle\|_{2,\infty} \|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$
$$\leq \sqrt{n_2 \ell} \|\mathcal{S}_\triangle\|_{2,\infty} \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|$$
$$\leq \sqrt{n_2 \ell} \sqrt{\alpha n_2 n_3} \|\mathcal{S}_\triangle\|_\infty \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|^2$$
$$\leq 2 n_2 \sqrt{\alpha n_3 \ell} \frac{\mu s_r}{n_3 \sqrt{n_1 n_2 \ell}} \bar{\sigma}_1(\mathcal{X}_\star) \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|^2$$
$$\leq 2 \frac{\mu s_r \kappa}{(1-\epsilon)^2} \sqrt{\frac{\alpha n_2}{n_1 n_3}} \bar{\sigma}_{s_r}(\mathcal{X}_\star) \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty}$$
$$\leq 2 \frac{\mu s_r \kappa}{(1-\epsilon)^2} \sqrt{\frac{\alpha n_2}{n_1 n_3}} \left(\sqrt{\frac{\mu s_r}{n_2 n_3 \ell}} + \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \right) \bar{\sigma}_{s_r}(\mathcal{X}_\star).$$

Note that $\|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \le \frac{\|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}}{\bar{\sigma}_{s_r}(\mathcal{X}_\star)}$. Thus,

$$
\sqrt{n_1}\|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \le \sqrt{\frac{\mu s_r}{n_3 \ell}}\Big(\frac{\epsilon}{1-\epsilon} + 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{n_3}}\Big)\bar{\sigma}_{s_r}(\mathcal{X}_\star)
$$
$$
+ 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{n_3}}\sqrt{n_2}\|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty},
$$

and similarly one can see

$$
\sqrt{n_2}\|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \le \sqrt{\frac{\mu s_r}{n_3 \ell}}\Big(\frac{\epsilon}{1-\epsilon} + 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{n_3}}\Big)\bar{\sigma}_{s_r}(\mathcal{X}_\star)
$$
$$
+ 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{n_3}}\sqrt{n_1}\|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}.
$$

Therefore,

$$
\sqrt{n_1}\|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \le \frac{\frac{\epsilon}{1-\epsilon} + 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{n_3}}}{1 - 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{n_3}}}\sqrt{\frac{\mu s_r}{n_3 \ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star)
$$
$$
\le \frac{\frac{\epsilon}{1-\epsilon} + 2\frac{c_0}{(1-\epsilon)^2}}{1 - 2\frac{c_0}{(1-\epsilon)^2}}\sqrt{\frac{\mu s_r}{n_3 \ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star)
$$
$$
= \frac{\frac{5c_0}{1-5c_0} + 2\frac{c_0}{(1-5c_0)^2}}{1 - 2\frac{c_0}{(1-5c_0)^2}}\sqrt{\frac{\mu s_r}{n_3 \ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star)
$$
$$
\le \sqrt{\frac{\mu s_r}{n_3 \ell}}\bar{\sigma}_{s_r}(\mathcal{X}_\star)
$$

as long as $c_0 \le 0.06$. This finishes the proof.

C.3 Proof of Theorem 13

Proof We set $c_0 \le 0.004$ in Lemma 21, then the results of Lemma 21 satisfy the condition of Lemma 20 and give

$$
\text{dist}(\mathcal{F}_t, \mathcal{F}_\star) \le \frac{0.02}{\sqrt{\ell}}(1 - 0.6\eta)^t \bar{\sigma}_{s_r}(\mathcal{X}_\star),
$$

$$
\sqrt{n_1}\|(\mathcal{L}_t *_\Phi \mathcal{Q}_t - \mathcal{L}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|(\mathcal{R}_t *_\Phi \mathcal{Q}_t^{-H} - \mathcal{R}_\star) *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \le \sqrt{\frac{\mu s_r}{n_3 \ell}}(1 - 0.6\eta)^t \bar{\sigma}_{s_r}(\mathcal{X}_\star)
$$

for all $t \ge 0$. According to Lemma 48, we have $\|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\| \vee \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\| \le \epsilon(1-0.6\eta)^t \le \epsilon$. According to (51), $\|\mathcal{L}_t *_\Phi \mathcal{R}_t^H - \mathcal{X}_\star\|_F \le 1.5\text{dist}(\mathcal{F}_t, \mathcal{F}_\star)$. Hence, the first claim is proved. The second and third claims are followed by Lemma 50 and Lemma 22, respectively. ∎

Appendix D. Proof for Tensor Completion

This section is devoted to the proofs of claims related to tensor completion.

D.1 Proof of Lemma 23

Lemma 51 (Tong et al. (2021a), Claim 5) *For tensor columns $\vec{\mathcal{A}}, \vec{\mathcal{A}}_\star \in \mathbb{R}^{n_1 \times 1 \times n_3}$ and $\lambda \geq \|\vec{\mathcal{A}}_\star\|_F / \|\vec{\mathcal{A}}\|_F$, it holds that*

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

We first prove the non-expansiveness property. Denote the optimal alignment tensor between $\widetilde{\mathcal{F}}$ and $\mathcal{F}_\star$ as $\widetilde{\mathcal{Q}}$, whose existence is guaranteed by Lemma 41. Let $\mathcal{P}_\varsigma(\widetilde{\mathcal{F}}) = \begin{bmatrix} \widetilde{\mathcal{L}} \\ \widetilde{\mathcal{R}} \end{bmatrix}$, by the definition of $\mathrm{dist}(\mathcal{P}_\varsigma(\widetilde{\mathcal{F}}), \mathcal{F}_\star)$, we have

$$\mathrm{dist}^2(\mathcal{P}_\varsigma(\widetilde{\mathcal{F}}), \mathcal{F}_\star) \leq \sum_{i=1}^{n_1} \|\mathcal{L}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} - (\mathcal{L}_\star *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}})(i,:,:)\|_F^2$$

$$+ \sum_{j=1}^{n_2} \|\mathcal{R}(j,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}}^{-H} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} - (\mathcal{R}_\star *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}})(j,:,:)\|_F^2. \quad (57)$$

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

$$\|(\widetilde{\mathcal{L}} *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}} - \mathcal{L}_\star) *_{\mathbf{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\| \vee \|(\widetilde{\mathcal{R}} *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}}^{-H} - \mathcal{R}_\star) *_{\mathbf{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\| \leq \epsilon.$$

Utilizing the fact that $\mathcal{R}_\star *_{\mathbf{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}} = \mathcal{V}_\star$, we arrive at

$$\|\widetilde{\mathcal{L}}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{R}}^H\|_F \leq \|\widetilde{\mathcal{L}}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\widetilde{\mathcal{R}} *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}}^{-H} *_{\mathbf{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\|$$

$$\leq \|\widetilde{\mathcal{L}}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \left(\|\mathcal{V}_\star\| + \|(\widetilde{\mathcal{R}} *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}}^{-H} - \mathcal{R}_\star) *_{\mathbf{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\|\right)$$

$$\leq (1+\epsilon)\|\widetilde{\mathcal{L}}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F.$$

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

$$\sqrt{n_1}\|(\mathcal{L}_\star *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}})(i,:,:)\|_F \leq \sqrt{n_1}\|\mathcal{U}_\star\|_{2,\infty}\|\mathcal{G}_\star\| \leq \sqrt{\frac{\mu s_r}{n_3 \ell}} \bar{\sigma}_1(\mathcal{X}_\star) \leq \frac{\varsigma}{1+\epsilon},$$

where the last inequality follows from the choice of ς. Take the above two inequalities to reach

$$\frac{\varsigma}{\sqrt{n_1}\|\widetilde{\mathcal{L}}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{R}}^H\|_F} \geq \frac{\|(\mathcal{L}_\star *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}})(i,:,:)\|_F}{\|\widetilde{\mathcal{L}}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F}.$$

We can then apply Lemma 51 with $\vec{\mathcal{A}} = \widetilde{\mathcal{L}}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}$, $\vec{\mathcal{A}}_\star = (\mathcal{L}_\star *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}})(i,:,:)$, and $\lambda = \frac{\varsigma}{\sqrt{n_1}\|\widetilde{\mathcal{L}}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{R}}^H\|_F}$ to obtain

$$\|\mathcal{L}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} - (\mathcal{L}_\star *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}})(i,:,:)\|_F^2$$

$$= \|\left(1 \wedge \frac{\varsigma}{\sqrt{n_1}\|\widetilde{\mathcal{L}}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{R}}^H\|_F}\right)\widetilde{\mathcal{L}}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} - (\mathcal{L}_\star *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}})(i,:,:)\|_F^2$$

$$\leq \|\widetilde{\mathcal{L}}(i,:,:) *_{\mathbf{\Phi}} \widetilde{\mathcal{Q}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} - (\mathcal{L}_\star *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}})(i,:,:)\|_F^2.$$

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

$$\mathrm{dist}^2(\mathcal{P}_\varsigma(\widetilde{\mathcal{F}}), \mathcal{F}_\star) \leq \sum_{i=1}^{n_1} \|\widetilde{\mathcal{L}}(i,:,:) *_{\boldsymbol{\Phi}} \widetilde{\mathcal{Q}} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} - (\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}})(i,:,:)\|_F^2$$

$$+ \sum_{j=1}^{n_2} \|\widetilde{\mathcal{R}}(j,:,:) *_{\boldsymbol{\Phi}} \widetilde{\mathcal{Q}}^{-H} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} - (\mathcal{R}_\star *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}})(j,:,:)\|_F^2 \leq \mathrm{dist}^2(\widetilde{\mathcal{F}}, \mathcal{F}_\star).$$

Next, we prove the incoherence condition. For any $i \in [n_1]$, one has

$$\|\mathcal{L}(i,:,:) *_{\boldsymbol{\Phi}} \mathcal{R}^H\|_F^2$$

$$= \sum_{j=1}^{n_2} \|\mathcal{L}(i,:,:) *_{\boldsymbol{\Phi}} \mathcal{R}(j,:,:)^H\|_F^2$$

$$= \sum_{j=1}^{n_2} \Big(1 \wedge \frac{\varsigma}{\sqrt{n_1}\|\widetilde{\mathcal{L}}(i,:,:) *_{\boldsymbol{\Phi}} \widetilde{\mathcal{R}}^H\|_F}\Big)^2 \|\widetilde{\mathcal{L}}(i,:,:) *_{\boldsymbol{\Phi}} \widetilde{\mathcal{R}}(j,:,:)^H\|_F^2 \Big(1 \wedge \frac{\varsigma}{\sqrt{n_2}\|\widetilde{\mathcal{R}}(j,:,:) *_{\boldsymbol{\Phi}} \widetilde{\mathcal{L}}^H\|_F}\Big)^2$$

$$\overset{(a)}{\leq} \Big(1 \wedge \frac{\varsigma}{\sqrt{n_1}\|\widetilde{\mathcal{L}}(i,:,:) *_{\boldsymbol{\Phi}} \widetilde{\mathcal{R}}^H\|_F}\Big)^2 \sum_{j=1}^{n_2} \|\widetilde{\mathcal{L}}(i,:,:) *_{\boldsymbol{\Phi}} \widetilde{\mathcal{R}}(j,:,:)^H\|_F^2$$

$$= \Big(1 \wedge \frac{\varsigma}{\sqrt{n_1}\|\widetilde{\mathcal{L}}(i,:,:) *_{\boldsymbol{\Phi}} \widetilde{\mathcal{R}}^H\|_F}\Big)^2 \|\widetilde{\mathcal{L}}(i,:,:) *_{\boldsymbol{\Phi}} \widetilde{\mathcal{R}}^H\|_F^2$$

$$\overset{(b)}{\leq} \frac{\varsigma^2}{n_1},$$

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

D.2 Proof of Lemma 24

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

Lemma 52 (Tropp (2012)) *Consider a finite sequence $\{\boldsymbol{Z}_k\}$ of independent, random $d_1 \times d_2$ matrices that satisfy the assumption $\mathbb{E}[\boldsymbol{Z}_k] = \boldsymbol{0}$ and $\|\boldsymbol{Z}_k\| \leq R$ almost surely. Let $\sigma^2 = \max\{\|\sum_k \mathbb{E}[\boldsymbol{Z}_k \boldsymbol{Z}_k^H]\|, \|\sum_k \mathbb{E}[\boldsymbol{Z}_k^H \boldsymbol{Z}_k]\|\}$. Then, for any $t \geq 0$, we have*

$$\mathbb{P}\Big[\|\sum_k \boldsymbol{Z}_k\| \geq t\Big] \leq (d_1 + d_2)\exp\Big(-\frac{t^2}{2\sigma^2 + \frac{2}{3}Rt}\Big)$$

$$\leq (d_1 + d_2)\exp\Big(-\frac{3t^2}{8\sigma^2}\Big), \quad \text{for } t \leq \sigma^2/R.$$

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

$$\|\sum_k \boldsymbol{Z}_k\| \leq 2\sqrt{c\sigma^2 \log(d_1 + d_2)} + cR\log(d_1 + d_2)$$

with probability at least $1 - (d_1 + d_2)^{1-c}$.

Lemma 53 *Suppose that $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is fixed, and $\boldsymbol{\Omega} \sim \mathrm{Ber}(p)$. Then with high probability,*

$$\|(p^{-1}\boldsymbol{\mathcal{P}_\Omega} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{Z}})\| \leq c\Big(\frac{\sqrt{\ell}\log((n_1 \vee n_2)n_3)}{p}\|\boldsymbol{\mathcal{Z}}\|_\infty + \sqrt{\frac{\ell\log((n_1 \vee n_2)n_3)}{p}}\|\boldsymbol{\mathcal{Z}}\|_{\infty,2}\Big),$$

for some numerical constant $c > 0$.

The following lemma establishes restricted strong convexity and smoothness of the observation operator for tensors in $\boldsymbol{T}$, which can be considered as an extension of Zheng and Lafferty (2016, Lemma 10).

Lemma 54 *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,*

$$p(1 - \epsilon)\|\boldsymbol{\mathcal{A}}\|_F^2 \leq \|\boldsymbol{\mathcal{P}_\Omega}(\boldsymbol{\mathcal{A}})\|_F^2 \leq p(1 + \epsilon)\|\boldsymbol{\mathcal{A}}\|_F^2. \tag{58}$$

Consequently,

$$|p^{-1}\langle\boldsymbol{\mathcal{P}_\Omega}(\boldsymbol{\mathcal{A}}), \boldsymbol{\mathcal{P}_\Omega}(\boldsymbol{\mathcal{B}})\rangle - \langle\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\rangle| \leq \epsilon\|\boldsymbol{\mathcal{A}}\|_F\|\boldsymbol{\mathcal{B}}\|_F, \tag{59}$$

provided that $p \geq c\epsilon^{-2}\mu s_r(n_1 + n_2)\log((n_1 \vee n_2)n_3)/(n_1 n_2 n_3)$ for some numerical constant $c > 0$.

We then have the following simple corollary.

Corollary 55 *Suppose that $\boldsymbol{\mathcal{X}}_\star$ is μ-incoherent, and $p \gtrsim \mu s_r(n_1+n_2)\log((n_1\vee n_2)n_3)/(n_1 n_2 n_3)$. Then with high probability,*

$$|\langle(p^{-1}\boldsymbol{\mathcal{P}_\Omega} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_A^H + \boldsymbol{\mathcal{L}}_A *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\star^H), \boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_B^H + \boldsymbol{\mathcal{L}}_B *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\star^H\rangle|$$

$$\leq c\sqrt{\frac{\mu s_r(n_1 + n_2)\log((n_1 \vee n_2)n_3)}{pn_1 n_2 n_3}}\|\boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_A^H + \boldsymbol{\mathcal{L}}_A *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\star^H\|_F\|\boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_B^H + \boldsymbol{\mathcal{L}}_B *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\star^H\|_F,$$

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

Lemma 56 *Suppose that $p \gtrsim \log((n_1 \vee n_2)n_3)/(n_1 \wedge n_2)$. Then with high probability,*

$$|\langle(p^{-1}\boldsymbol{\mathcal{P}_\Omega} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_A *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_A^H), \boldsymbol{\mathcal{L}}_B *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_B^H\rangle|$$

$$\leq c\ell^{\frac{3}{2}}\sqrt{\frac{(n_1 \vee n_2)\log((n_1 \vee n_2)n_3)}{p}}$$

$$\Big(\|\boldsymbol{\mathcal{L}}_A\|_{2,2,\infty}\|\boldsymbol{\mathcal{L}}_B\|_F \wedge \|\boldsymbol{\mathcal{L}}_A\|_F\|\boldsymbol{\mathcal{L}}_B\|_{2,2,\infty}\Big)\Big(\|\boldsymbol{\mathcal{R}}_A\|_{2,2,\infty}\|\boldsymbol{\mathcal{R}}_B\|_F \wedge \|\boldsymbol{\mathcal{R}}_A\|_F\|\boldsymbol{\mathcal{R}}_B\|_{2,2,\infty}\Big),$$

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

Lemma 57 *Under conditions* $\mathrm{dist}(\mathcal{F}_t, \mathcal{F}_\star) \leq \frac{\epsilon}{\sqrt{\ell}} \bar{\sigma}_{s_r}(\mathcal{X}_\star)$ *and* $\sqrt{n_1} \|\mathcal{L}_\sharp *_\Phi \mathcal{R}_\sharp^H\|_{2,\infty} \vee \sqrt{n_2} \|\mathcal{R}_\sharp *_\Phi$
$\mathcal{L}_\sharp^H\|_{2,\infty} \leq c_\varsigma \sqrt{\frac{\mu s_r}{n_3 \ell}} \bar{\sigma}_1(\mathcal{X}_\star)$, *we have*

$$\|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\| \vee \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\| \leq \epsilon; \tag{60a}$$

$$\|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \leq \frac{1}{1-\epsilon}; \tag{60b}$$

$$\|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \leq \frac{1}{(1-\epsilon)^2}; \tag{60c}$$

$$\sqrt{n_1} \|\mathcal{L}_\sharp *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2} \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \frac{c_\varsigma}{1-\epsilon} \sqrt{\frac{\mu s_r}{n_3 \ell}} \bar{\sigma}_1(\mathcal{X}_\star); \tag{60d}$$

$$\sqrt{n_1} \|\mathcal{L}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2} \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \leq \frac{c_\varsigma \kappa}{1-\epsilon} \sqrt{\frac{\mu s_r}{n_3 \ell}}; \tag{60e}$$

$$\sqrt{n_1} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2} \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \leq (1 + \frac{c_\varsigma}{1-\epsilon}) \sqrt{\frac{\mu s_r}{n_3 \ell}} \bar{\sigma}_1(\mathcal{X}_\star). \tag{60f}$$

Proof First, repeating the derivation for Lemma 41 obtains (60a). Second, taking the condition (60a) and Lemma 44 together to obtain (60b) and (60c). Third, taking the incoherence condition $\sqrt{n_1} \|\mathcal{L}_\sharp *_\Phi \mathcal{R}_\sharp^H\|_{2,\infty} \vee \sqrt{n_2} \|\mathcal{R}_\sharp *_\Phi \mathcal{L}_\sharp^H\|_{2,\infty} \leq c_\varsigma \sqrt{\frac{\mu s_r}{n_3 \ell}} \bar{\sigma}_1(\mathcal{X}_\star)$, and $\|\mathcal{A} *_\Phi \mathcal{B}\|_{2,\infty} \geq \|\mathcal{A}\|_{2,\infty} \bar{\sigma}_{s_r}(\mathcal{B})$ from Lemma 36, together with the relations

$$\|\mathcal{L}_\sharp *_\Phi \mathcal{R}_\sharp^H\|_{2,\infty} \geq \bar{\sigma}_{s_r}(\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}) \|\mathcal{L}_\sharp *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$

$$\geq (\bar{\sigma}_{s_r}(\mathcal{R}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}) - \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|) \|\mathcal{L}_\sharp *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$

$$\geq (1-\epsilon) \|\mathcal{L}_\sharp *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty};$$

$$\|\mathcal{R}_\sharp *_\Phi \mathcal{L}_\sharp^H\|_{2,\infty} \geq \bar{\sigma}_{s_r}(\mathcal{L}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}) \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$

$$\geq (\bar{\sigma}_{s_r}(\mathcal{L}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}) - \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|) \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$

$$\geq (1-\epsilon) \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$

to obtain (60d) and (60e). Finally, (60f) can be obtained by applying the triangle inequality together with incoherence assumption. ∎

Now we prove Lemma 24. First, we define the event G as the intersection of the events that the bounds in Corollary 55 and Lemma 56 hold. The rest of the proof is under the assumption that G holds, which happens with high probability. By the condition $\mathrm{dist}(\mathcal{F}_t, \mathcal{F}_\star) \leq \frac{0.02}{\sqrt{\ell}} \bar{\sigma}_{s_r}(\mathcal{X}_\star)$ and Lemma 41, one knows that $\mathcal{Q}_t$, the optimal alignment tensor between $\mathcal{F}_t$ and $\mathcal{F}_\star$ exists, and $\epsilon := 0.02$. In addition, denote $\widetilde{\mathcal{F}}_{t+1}$ as the update before projection as

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

and therefore $\boldsymbol{\mathcal{F}}_{t+1} = \mathcal{P}_\varsigma(\widetilde{\boldsymbol{\mathcal{F}}}_{t+1})$. Note that it suffices to prove the following relation

$$\text{dist}(\widetilde{\boldsymbol{\mathcal{F}}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) \leq (1 - 0.6\eta)\text{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star), \tag{61}$$

since the conclusion $\|\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H - \boldsymbol{\mathcal{X}}_\star\|_F \leq 1.5\text{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star)$ is a simple consequence of Lemma 45; see (51) for details. In the following, we focus on proving (61). By the definition of $\text{dist}(\widetilde{\boldsymbol{\mathcal{F}}}_{t+1}, \boldsymbol{\mathcal{F}}_\star)$, we have

$$\text{dist}^2(\widetilde{\boldsymbol{\mathcal{F}}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) \leq \|(\widetilde{\boldsymbol{\mathcal{L}}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \|(\widetilde{\boldsymbol{\mathcal{R}}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2. \tag{62}$$

Plugging in the update rule (21) and the decomposition $\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H + \boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\triangle^H$ to obtain

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

This allows us to expand the square of the first term in (62) as

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

Bound of $\mathfrak{P}_1$. The first term $\mathfrak{P}_1$ has already been controlled in (53) as follows.

$$\mathfrak{P}_1 \leq \left((1 - \eta)^2 + \frac{2\epsilon}{1 - \epsilon}\eta(1 - \eta)\right)\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \frac{2\epsilon + \epsilon^2}{(1 - \epsilon)^2}\eta^2\|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2.$$

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

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

For the first term $\mathfrak{P}_{2,1}$, we can invoke Corollary 55 to obtain

$$
\begin{aligned}
\mathfrak{P}_{2,1} \leq {}& c_1\sqrt{\frac{\mu s_r(n_1+n_2)\log((n_1 \vee n_2)n_3)}{pn_1n_2n_3}}\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{R}_\star^H\|_F\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{R}_\star^H\|_F \\
= {}& c_1\sqrt{\frac{\mu s_r(n_1+n_2)\log((n_1 \vee n_2)n_3)}{pn_1n_2n_3}}\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}} *_{\mathbf{\Phi}} \mathcal{R}_\star^H\|_F \\
& \|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} *_{\mathbf{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}} *_{\mathbf{\Phi}} \mathcal{R}_\star^H\|_F \\
\leq {}& c_1\sqrt{\frac{\mu s_r(n_1+n_2)\log((n_1 \vee n_2)n_3)}{pn_1n_2n_3}}\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\|\mathcal{G}_\star^{-\frac{1}{2}} *_{\mathbf{\Phi}} \mathcal{R}_\star^H\| \\
& \|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\|\mathcal{G}_\star^{\frac{1}{2}} *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|\|\mathcal{G}_\star^{-\frac{1}{2}} *_{\mathbf{\Phi}} \mathcal{R}_\star^H\| \\
= {}& c_1\sqrt{\frac{\mu s_r(n_1+n_2)\log((n_1 \vee n_2)n_3)}{pn_1n_2n_3}}\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2\|\mathcal{G}_\star^{\frac{1}{2}} *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \\
\leq {}& \frac{c_1}{(1-\epsilon)^2}\sqrt{\frac{\mu s_r(n_1+n_2)\log((n_1 \vee n_2)n_3)}{pn_1n_2n_3}}\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2,
\end{aligned}
$$

where the last inequality uses (60c). For the term $\mathfrak{P}_{2,2}$, we can invoke Lemma 56 with $\mathcal{L}_A := \mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}$, $\mathcal{R}_A := \mathcal{R}_\star *_{\mathbf{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}$, $\mathcal{L}_B := \mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}$, $\mathcal{R}_B := \mathcal{R}_\triangle *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}$

to obtain

$$\mathfrak{P}_{2,2} \le c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2) \log((n_1 \vee n_2)n_3)}{p}} \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,2,\infty} \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F$$

$$\|\boldsymbol{\mathcal{R}}_{\star} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_{2,2,\infty} \|\boldsymbol{\mathcal{R}}_{\triangle} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_{\sharp}^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_{\sharp})^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F$$

$$\le c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2) \log((n_1 \vee n_2)n_3)}{p}} \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F$$

$$\|\boldsymbol{\mathcal{R}}_{\star} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{R}}_{\triangle} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_{\sharp}^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_{\sharp})^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F$$

$$\le c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2) \log((n_1 \vee n_2)n_3)}{p}} \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F$$

$$\|\boldsymbol{\mathcal{R}}_{\star} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{R}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_{\sharp}^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_{\sharp})^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|.$$

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

$$\mathfrak{P}_{2,3} \le c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2) \log((n_1 \vee n_2)n_3)}{p}} \|\boldsymbol{\mathcal{L}}_{\sharp} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F$$

$$\|\boldsymbol{\mathcal{R}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{R}}_{\sharp} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_{2,\infty} \|\boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_{\sharp}^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_{\sharp})^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|.$$

Utilizing the consequences in Lemma 57, we have

$$\mathfrak{P}_{2,2} \le \frac{c_2}{(1-\epsilon)^2} \Big(1 + \frac{c_\varsigma}{1-\epsilon}\Big) \sqrt{\frac{\ell \log((n_1 \vee n_2)n_3)}{p(n_1 \wedge n_2)}} \frac{\mu s_r \kappa}{n_3} \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{R}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F;$$

$$\mathfrak{P}_{2,3} \le \frac{c_2}{(1-\epsilon)^4} \sqrt{\frac{\ell \log((n_1 \vee n_2)n_3)}{p(n_1 \wedge n_2)}} \frac{\mu s_r c_\varsigma^2 \kappa^2}{n_3} \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{R}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F.$$

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

$$|\mathfrak{P}_2| \le 2\eta(1-\eta) \Big(\frac{c_1}{(1-\epsilon)^2} \sqrt{\frac{\mu s_r (n_1 + n_2) \log((n_1 \vee n_2)n_3)}{p n_1 n_2 n_3}} \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F^2$$

$$+ \frac{c_2}{(1-\epsilon)^2} \Big(1 + \frac{c_\varsigma}{1-\epsilon} + \frac{c_\varsigma^2 \kappa}{(1-\epsilon)^2}\Big) \sqrt{\frac{\ell \log((n_1 \vee n_2)n_3)}{p(n_1 \wedge n_2)}} \frac{\mu s_r \kappa}{n_3} \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{R}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F\Big)$$

$$= 2\eta(1-\eta) \Big(\nu_1 \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F^2 + \nu_2 \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F \|\boldsymbol{\mathcal{R}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F\Big)$$

$$\le \eta(1-\eta) \Big((2\nu_1 + \nu_2) \|\boldsymbol{\mathcal{L}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F^2 + \nu_2 \|\boldsymbol{\mathcal{R}}_{\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_F^2\Big),$$

where we denote

$$\nu_1 := \frac{c_1}{(1-\epsilon)^2} \sqrt{\frac{\mu s_r (n_1 + n_2) \log((n_1 \vee n_2)n_3)}{p n_1 n_2 n_3}}$$

$$\text{and} \quad \nu_2 := \frac{c_2}{(1-\epsilon)^2} \Big(1 + \frac{c_\varsigma}{1-\epsilon} + \frac{c_\varsigma^2 \kappa}{(1-\epsilon)^2}\Big) \sqrt{\frac{\ell \log((n_1 \vee n_2)n_3)}{p(n_1 \wedge n_2)}} \frac{\mu s_r \kappa}{n_3}. \tag{63}$$

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

$$
\begin{aligned}
|\mathfrak{P}_3| \leq 2\eta^2 \Big(\Big| \langle \mathcal{L}_\star *_\Phi \mathcal{R}_\triangle^H *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}, \\
(p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\triangle *_\Phi \mathcal{R}_\sharp^H) *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}} \rangle \Big| \\
+ \Big| \langle \mathcal{L}_\star *_\Phi \mathcal{R}_\triangle^H *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}, \\
(p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\star *_\Phi \mathcal{R}_\triangle^H) *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}} \rangle \Big| \Big) \\
:= 2\eta^2 (\mathfrak{P}_{3,1} + \mathfrak{P}_{3,2}).
\end{aligned}
$$

For $\mathfrak{P}_{3,1}$, we invoke Lemma 56 with $\mathcal{L}_A := \mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}$, $\mathcal{R}_A := \mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}$, $\mathcal{L}_B := \mathcal{L}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}$, $\mathcal{R}_B := \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}$ and use the consequences in Lemma 57 to obtain

$$
\begin{aligned}
\mathfrak{P}_{3,1} &\leq c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2)\log((n_1 \vee n_2)n_3)}{p}} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{L}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,2,\infty} \\
&\quad \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,2,\infty} \|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \\
&\leq c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2)\log((n_1 \vee n_2)n_3)}{p}} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{L}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \\
&\quad \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \\
&\leq c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2)\log((n_1 \vee n_2)n_3)}{p}} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{L}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \\
&\quad \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|^2 \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \\
&\leq \frac{c_2}{(1-\epsilon)^3} \sqrt{\frac{\ell \log((n_1 \vee n_2)n_3)}{p(n_1 \wedge n_2)}} \frac{\mu s_r c_\varsigma \kappa}{n_3} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F.
\end{aligned}
$$

For $\mathfrak{P}_{3,2}$, we again invoke Corollary 55 to obtain

$$
\begin{aligned}
\mathfrak{P}_{3,2} &\leq c_1 \sqrt{\frac{\mu s_r (n_1 + n_2)\log((n_1 \vee n_2)n_3)}{pn_1 n_2 n_3}} \|\mathcal{L}_\star *_\Phi \mathcal{R}_\triangle^H\|_F \\
&\quad \|\mathcal{L}_\star *_\Phi \mathcal{R}_\triangle^H *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\sharp^H\|_F \\
&\leq c_1 \sqrt{\frac{\mu s_r (n_1 + n_2)\log((n_1 \vee n_2)n_3)}{pn_1 n_2 n_3}} \|\mathcal{L}_\star *_\Phi \mathcal{R}_\triangle^H\|_F^2 \|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|^2 \\
&\leq \frac{c_1}{(1-\epsilon)^2} \sqrt{\frac{\mu s_r (n_1 + n_2)\log((n_1 \vee n_2)n_3)}{pn_1 n_2 n_3}} \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F^2,
\end{aligned}
$$

61

where the last inequality uses (60b). Combining the bounds for $\mathfrak{P}_{3,1}$ and $\mathfrak{P}_{3,2}$ to arrive at

$$|\mathfrak{P}_3| \le 2\eta^2 \Big(\frac{c_1}{(1-\epsilon)^2} \sqrt{\frac{\mu s_r(n_1+n_2)\log((n_1 \vee n_2)n_3)}{pn_1n_2n_3}} \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2$$

$$+ \frac{c_2}{(1-\epsilon)^3} \sqrt{\frac{\ell \log((n_1 \vee n_2)n_3)}{p(n_1 \wedge n_2)}} \frac{\mu s_r c_\varsigma \kappa}{n_3} \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \Big)$$

$$\le 2\eta^2 \Big(\nu_1 \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \nu_2 \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \Big)$$

$$\le \eta^2 \Big(\nu_2 \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + (2\nu_1+\nu_2) \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 \Big).$$

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

$$\sqrt{\mathfrak{P}_4} = \eta \|(p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F$$

$$= \eta \max_{\widetilde{\mathcal{L}} \in \mathbb{R}^{n_1 \times r \times n_3} : \|\widetilde{\mathcal{L}}\|_F \le 1} \langle (p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}, \widetilde{\mathcal{L}} \rangle$$

$$= \eta \max_{\widetilde{\mathcal{L}} \in \mathbb{R}^{n_1 \times r \times n_3} : \|\widetilde{\mathcal{L}}\|_F \le 1} \langle (p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star), \widetilde{\mathcal{L}} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H \rangle$$

$$\le \eta \Big(\Big| \langle (p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H), \widetilde{\mathcal{L}} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H \rangle \Big|$$

$$+ \Big| \langle (p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H), \widetilde{\mathcal{L}} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H \rangle \Big|$$

$$+ \Big| \langle (p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H), \widetilde{\mathcal{L}} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H \rangle \Big|$$

$$:= \eta(\mathfrak{P}_{4,1} + \mathfrak{P}_{4,2} + \mathfrak{P}_{4,3}).$$

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

$$\mathfrak{P}_{4,1} \le \frac{c_1}{(1-\epsilon)^2} \sqrt{\frac{\mu s_r(n_1+n_2)\log((n_1 \vee n_2)n_3)}{pn_1n_2n_3}} \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F;$$

$$\mathfrak{P}_{4,2} \le \frac{c_2}{(1-\epsilon)^2} \Big(1 + \frac{c_\varsigma}{1-\epsilon}\Big) \sqrt{\frac{\ell \log((n_1 \vee n_2)n_3)}{p(n_1 \wedge n_2)}} \frac{\mu s_r \kappa}{n_3} \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F;$$

$$\mathfrak{P}_{4,3} \le \frac{c_2}{(1-\epsilon)^4} \sqrt{\frac{\ell \log((n_1 \vee n_2)n_3)}{p(n_1 \wedge n_2)}} \frac{\mu s_r c_\varsigma^2 \kappa^2}{n_3} \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F.$$

Hence,

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

Taking the square on both sides to obtain the upper bound

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

Taking the bounds for $\mathfrak{P}_1, \mathfrak{P}_2, \mathfrak{P}_3$ and $\mathfrak{P}_4$ collectively yields

$$
\begin{aligned}
&\|(\widetilde{\boldsymbol{\mathcal{L}}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 \\
&\leq \Big((1-\eta)^2 + \frac{2\epsilon}{1-\epsilon}\eta(1-\eta) \Big) \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \frac{2\epsilon+\epsilon^2}{(1-\epsilon)^2}\eta^2 \|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 \\
&\quad + \eta(1-\eta) \Big((2\nu_1+\nu_2)\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \nu_2 \|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 \Big) \\
&\quad + \eta^2 \Big(\nu_2 \|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + (2\nu_1+\nu_2)\|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 \Big) \\
&\quad + \eta^2 \Big(\nu_1(\nu_1+\nu_2)\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \nu_2(\nu_1+\nu_2)\|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 \Big).
\end{aligned}
$$

A similar upper bound also holds for the second term in (62). It then turns out that

$$
\|(\widetilde{\boldsymbol{\mathcal{L}}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \|(\widetilde{\boldsymbol{\mathcal{R}}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 \leq \hbar^2(\eta;\epsilon,\nu_1,\nu_2)\mathrm{dist}^2(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star),
$$

where the contraction rate $\hbar^2(\eta;\epsilon,\nu_1,\nu_2)$ is given by

$$
\hbar^2(\eta;\epsilon,\nu_1,\nu_2) := (1-\eta)^2 + \Big(\frac{2\epsilon}{1-\epsilon} + 2(\nu_1+\nu_2) \Big)\eta(1-\eta) + \Big(\frac{2\epsilon+\epsilon^2}{(1-\epsilon)^2} + 2(\nu_1+\nu_2) + (\nu_1+\nu_2)^2 \Big)\eta^2.
$$

As long as $p \geq c\Big(\frac{\mu s_r(n_1+n_2)\log((n_1\vee n_2)n_3)}{n_1 n_2 n_3} \vee \frac{\mu^2 s_r^2 \kappa^4 \ell \log((n_1\vee n_2)n_3)}{(n_1\wedge n_2)n_3^2} \Big)$ for some sufficiently large constant c, we have $\nu_1 + \nu_2 \leq 0.1$ under the setting $\epsilon = 0.02$. When $0 < \eta \leq \frac{2}{3}$, we further have $\hbar(\eta;\epsilon,\nu_1,\nu_2) \leq 1 - 0.6\eta$. Thus we conclude that

$$
\begin{aligned}
\mathrm{dist}(\widetilde{\boldsymbol{\mathcal{F}}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) &\leq \sqrt{\|(\widetilde{\boldsymbol{\mathcal{L}}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \|(\widetilde{\boldsymbol{\mathcal{R}}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2} \\
&\leq (1 - 0.6\eta)\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star).
\end{aligned}
$$

This finishes the proof.

D.2.1 PROOF OF LEMMA 53

Denote the tensor $\boldsymbol{\mathcal{H}}_{ijk} = (p^{-1}\delta_{ijk} - 1)\boldsymbol{\mathcal{Z}}_{i,j,k}\bar{\boldsymbol{\mathfrak{e}}}_{ijk}$. Then we have

$$
(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{Z}}) = \sum_{i,j,k} \boldsymbol{\mathcal{H}}_{ijk}.
$$

Note that δ_{ijk}'s are independent random scalars. Thus, $\boldsymbol{\mathcal{H}}_{ijk}$'s are independent random tensors and $\bar{\boldsymbol{H}}_{ijk}$'s are independent random matrices. Observe that $\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}] = \boldsymbol{0}$ and $\|\bar{\boldsymbol{H}}_{ijk}\| \leq p^{-1}\sqrt{\ell}\|\boldsymbol{\mathcal{Z}}\|_\infty$ by using (35). According to the proof in Lemma 38, we know that $\sum_{i,j,k} \boldsymbol{\mathcal{Z}}_{i,j,k}^2 L(\bar{\boldsymbol{\mathfrak{e}}}_{ijk}^H *_{\boldsymbol{\Phi}} \bar{\boldsymbol{\mathfrak{e}}}_{ijk})$ is f-diagonal and the k'-th entry of its (j,j)-th mode-3 tube is $\sum_{i,k} \boldsymbol{\mathcal{Z}}_{i,j,k}^2 |\boldsymbol{\Phi}_{k',k}|^2$. We have

$$
\sum_{i,k} \boldsymbol{\mathcal{Z}}_{i,j,k}^2 |\boldsymbol{\Phi}_{k',k}|^2 \leq \sum_i \Big(\sum_k \boldsymbol{\mathcal{Z}}_{i,j,k}^2 \Big)\Big(\max_k |\boldsymbol{\Phi}_{k',k}|^2 \Big) \leq \sum_{i,k} \boldsymbol{\mathcal{Z}}_{i,j,k}^2 \|\boldsymbol{\Phi}\|_\infty^2 \leq \ell \sum_{i,k} \boldsymbol{\mathcal{Z}}_{i,j,k}^2.
$$

Hence, the tensor spectral norm of $\sum_{i,j,k} \boldsymbol{\mathcal{Z}}_{i,j,k}^2 (\bar{\mathbf{e}}_{ijk}^H *_{\boldsymbol{\Phi}} \bar{\mathbf{e}}_{ijk})$ has the bound

$$\Big\| \sum_{i,j,k} \boldsymbol{\mathcal{Z}}_{i,j,k}^2 (\bar{\mathbf{e}}_{ijk}^H *_{\boldsymbol{\Phi}} \bar{\mathbf{e}}_{ijk}) \Big\| \le \max_j \ell \sum_{i,k} \boldsymbol{\mathcal{Z}}_{i,j,k}^2 \le \ell \|\boldsymbol{\mathcal{Z}}\|_{\infty,2}^2.$$

Then we have

$$
\begin{aligned}
\Big\| \sum_{i,j,k} \mathbb{E}[\bar{\boldsymbol{H}}_{ijk}^H \bar{\boldsymbol{H}}_{ijk}] \Big\| &= \Big\| \sum_{i,j,k} \mathbb{E}[\boldsymbol{\mathcal{H}}_{ijk}^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{H}}_{ijk}] \Big\| \\
&= \Big\| \sum_{i,j,k} \mathbb{E}[(1 - p^{-1}\delta_{ijk})^2] \boldsymbol{\mathcal{Z}}_{i,j,k} (\bar{\mathbf{e}}_{ijk}^H *_{\boldsymbol{\Phi}} \bar{\mathbf{e}}_{ijk}) \Big\| \\
&\le \frac{(1-p)\ell}{p} \|\boldsymbol{\mathcal{Z}}\|_{\infty,2}^2 \\
&\le p^{-1}\ell \|\boldsymbol{\mathcal{Z}}\|_{\infty,2}^2.
\end{aligned}
$$

We can also have a similar calculation that yields $\| \sum_{i,j,k} \mathbb{E}[\bar{\boldsymbol{H}}_{ijk} \bar{\boldsymbol{H}}_{ijk}^H] \| \le p^{-1}\ell \|\boldsymbol{\mathcal{Z}}\|_{\infty,2}^2$. The proof is completed by applying the matrix Bernstein inequality in Lemma 52.

D.2.2 Proof of Lemma 54

By Lemma 40, with high probability, for any $\boldsymbol{\mathcal{A}} \in \boldsymbol{T}$, it holds that

$$p(1-\epsilon)\|\boldsymbol{\mathcal{A}}\|_F \le \|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} \boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}})\|_F \le p(1+\epsilon)\|\boldsymbol{\mathcal{A}}\|_F. \tag{64}$$

Rewriting $\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{A}})\|_F = \langle \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} \boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}}), \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} \boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}}) \rangle = \langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{P}}_{\boldsymbol{T}} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} \boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}}) \rangle$, and using the Cauchy-Schwarz inequality, we can bound

$$\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{A}})\|_F^2 \le p(1+\epsilon)\|\boldsymbol{\mathcal{A}}\|_F^2. \tag{65}$$

In addition, we have

$$
\begin{aligned}
\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{A}})\|_F^2 &= \langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{P}}_{\boldsymbol{T}} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} \boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}}) \rangle = \langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{P}}_{\boldsymbol{T}} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} \boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}}) - p\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}}) + p\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}}) \rangle \\
&\ge -\|\boldsymbol{\mathcal{A}}\|_F \|\boldsymbol{\mathcal{P}}_{\boldsymbol{T}} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} \boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}}) - p\boldsymbol{\mathcal{P}}_{\boldsymbol{T}}(\boldsymbol{\mathcal{A}})\|_F + p\|\boldsymbol{\mathcal{A}}\|_F^2 \\
&\ge p(1-\epsilon)\|\boldsymbol{\mathcal{A}}\|_F^2, \tag{66}
\end{aligned}
$$

where the last inequality follows from Lemma 40. Combining (65) and (66) proves (58). To show (59), let $\boldsymbol{\mathcal{A}}' = \frac{\boldsymbol{\mathcal{A}}}{\|\boldsymbol{\mathcal{A}}\|_F}$ and $\boldsymbol{\mathcal{B}}' = \frac{\boldsymbol{\mathcal{B}}}{\|\boldsymbol{\mathcal{B}}\|_F}$. Both $\boldsymbol{\mathcal{A}}' + \boldsymbol{\mathcal{B}}'$ and $\boldsymbol{\mathcal{A}}' - \boldsymbol{\mathcal{B}}'$ are in $\boldsymbol{T}$. We have

$$
\begin{aligned}
\langle \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{A}}'), \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{B}}') \rangle &= \frac{1}{4}\Big(\|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{A}}' + \boldsymbol{\mathcal{B}}')\|_F^2 - \|\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{A}}' - \boldsymbol{\mathcal{B}}')\|_F^2 \Big) \\
&\le \frac{1}{4}\Big(p(1+\epsilon)\|\boldsymbol{\mathcal{A}}' + \boldsymbol{\mathcal{B}}'\|_F^2 - p(1-\epsilon)\|\boldsymbol{\mathcal{A}}' - \boldsymbol{\mathcal{B}}'\|_F^2 \Big) \\
&= \frac{1}{4}\Big(2p\epsilon(\|\boldsymbol{\mathcal{A}}'\|_F^2 + \|\boldsymbol{\mathcal{B}}'\|_F^2) + 4p\langle \boldsymbol{\mathcal{A}}', \boldsymbol{\mathcal{B}}' \rangle \Big) \\
&= p\epsilon + p\langle \boldsymbol{\mathcal{A}}', \boldsymbol{\mathcal{B}}' \rangle,
\end{aligned}
$$

where the first inequality follows from (58). Thus, we have

$$p^{-1}\langle \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{A}}), \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{B}}) \rangle = p^{-1}\|\boldsymbol{\mathcal{A}}\|_F \|\boldsymbol{\mathcal{B}}\|_F \langle \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{A}}'), \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{B}}') \rangle \le \epsilon \|\boldsymbol{\mathcal{A}}\|_F \|\boldsymbol{\mathcal{B}}\|_F + \langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle.$$

Similarly, we can show

$$p^{-1}\langle \mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{A}), \mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{B})\rangle \geq -\epsilon\|\mathcal{A}\|_F\|\mathcal{B}\|_F + \langle \mathcal{A}, \mathcal{B}\rangle.$$

The proof is completed.

D.2.3 PROOF OF LEMMA 56

First, using the definition of tensor inner product, we have

$$|\langle \mathcal{A}, \mathcal{B}\rangle| = \frac{1}{\ell}|\langle \bar{\boldsymbol{A}}, \bar{\boldsymbol{B}}\rangle| \leq \frac{1}{\ell}\|\bar{\boldsymbol{A}}\|\|\bar{\boldsymbol{B}}\|_* = \|\mathcal{A}\|\|\mathcal{B}\|_*,$$

for any $\mathcal{A}, \mathcal{B}$, where the inequality holds by matrix Hölder's inequality. Hence,

$$\begin{aligned}
&|p^{-1}\langle \mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{L}_A *_{\boldsymbol{\Phi}} \mathcal{R}_A^H), \mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{L}_B *_{\boldsymbol{\Phi}} \mathcal{R}_B^H)\rangle - \langle \mathcal{L}_A *_{\boldsymbol{\Phi}} \mathcal{R}_A^H, \mathcal{L}_B *_{\boldsymbol{\Phi}} \mathcal{R}_B^H\rangle| \\
=&|\langle (p^{-1}\mathcal{P}_{\boldsymbol{\Omega}} - \mathcal{I}_{n_1})(\mathcal{J}), ((\mathcal{L}_A *_{\boldsymbol{\Phi}} \mathcal{R}_A^H) \circ (\mathcal{L}_B *_{\boldsymbol{\Phi}} \mathcal{R}_B^H))\rangle| \\
\leq&\|(p^{-1}\mathcal{P}_{\boldsymbol{\Omega}} - \mathcal{I}_{n_1})(\mathcal{J})\|\|(\mathcal{L}_A *_{\boldsymbol{\Phi}} \mathcal{R}_A^H) \circ (\mathcal{L}_B *_{\boldsymbol{\Phi}} \mathcal{R}_B^H)\|_*,
\end{aligned} \tag{67}$$

where $\mathcal{J}$ denotes the tensor with all-one entries and $\circ$ denotes the Hadamard (elementwise) product. To bound $\|(p^{-1}\mathcal{P}_{\boldsymbol{\Omega}} - \mathcal{I}_{n_1})(\mathcal{J})\|$, we again denote the tensor $\mathcal{H}_{ijk} = (p^{-1}\delta_{ijk} - 1)\bar{\mathbf{e}}_{ijk}$. Then we have

$$(p^{-1}\mathcal{P}_{\boldsymbol{\Omega}} - \mathcal{I}_{n_1})(\mathcal{J}) = \sum_{i,j,k} \mathcal{H}_{ijk}.$$

Note that δ_{ijk}'s are independent random scalars. Thus, $\mathcal{H}_{ijk}$'s are independent random tensors and $\bar{\boldsymbol{H}}_{ijk}$'s are independent random matrices. Observe that $\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}] = \mathbf{0}$ and $\|\bar{\boldsymbol{H}}_{ijk}\| \leq p^{-1}\sqrt{\ell}$. Then we have

$$\begin{aligned}
\Big\|\sum_{i,j,k} \mathbb{E}[\bar{\boldsymbol{H}}_{ijk}^H \bar{\boldsymbol{H}}_{ijk}]\Big\| &= \Big\|\sum_{i,j,k} \mathbb{E}[(1 - p^{-1}\delta_{ijk})^2](\bar{\mathbf{e}}_{ijk}^H *_{\boldsymbol{\Phi}} \bar{\mathbf{e}}_{ijk})\Big\| \\
&= \Big\|\frac{1-p}{p}\sum_{i,j,k}(\bar{\mathbf{e}}_{ijk}^H *_{\boldsymbol{\Phi}} \bar{\mathbf{e}}_{ijk})\Big\| \\
&\leq \frac{1-p}{p}\Big\|\sum_{i,j,k}(\bar{\mathbf{e}}_{ijk}^H *_{\boldsymbol{\Phi}} \bar{\mathbf{e}}_{ijk})\Big\| \\
&\leq \frac{n_1\ell}{p},
\end{aligned}$$

where the last step uses (36). A similar calculation yields $\|\sum_{i,j,k} \mathbb{E}[\bar{\boldsymbol{H}}_{ijk}\bar{\boldsymbol{H}}_{ijk}^H]\| \leq \frac{n_2\ell}{p}$. Using Lemma 52, let $\epsilon = \sqrt{c(n_1 \vee n_2)\ell \log((n_1 \vee n_2)n_3)/p}$, then with high probability,

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

provided that $p \geq c\log((n_1 \vee n_2)n_3)/(n_1 \wedge n_2)$. To give a bound of $\|(\mathcal{L}_A *_{\boldsymbol{\Phi}} \mathcal{R}_A^H) \circ (\mathcal{L}_B *_{\boldsymbol{\Phi}} \mathcal{R}_B^H)\|_*$, notice that it is equal to

$$\|(\mathcal{L}_A *_{\boldsymbol{\Phi}} \mathcal{R}_A^H) \circ (\mathcal{L}_B *_{\boldsymbol{\Phi}} \mathcal{R}_B^H)\|_* = \sum_{k=1}^{n_3} \frac{1}{\ell}\|(\bar{\mathcal{L}}_A(:,:,k)\bar{\mathcal{R}}_A(:,:,k)^H) \circ (\bar{\mathcal{L}}_B(:,:,k)\bar{\mathcal{R}}_B(:,:,k)^H)\|_*,$$

and we can decompose each $(\bar{\boldsymbol{\mathcal{L}}}_A(:,:,k)\bar{\boldsymbol{\mathcal{R}}}_A(:,:,k)^H) \circ (\bar{\boldsymbol{\mathcal{L}}}_B(:,:,k)\bar{\boldsymbol{\mathcal{R}}}_B(:,:,k)^H)$ into sum of rank one matrices as follows

$$
\begin{aligned}
&(\bar{\boldsymbol{\mathcal{L}}}_A(:,:,k)\bar{\boldsymbol{\mathcal{R}}}_A(:,:,k)^H) \circ (\bar{\boldsymbol{\mathcal{L}}}_B(:,:,k)\bar{\boldsymbol{\mathcal{R}}}_B(:,:,k)^H) \\
&= \Big(\sum_{j=1}^{r_1} \bar{\boldsymbol{\mathcal{L}}}_A(:,j,k)\bar{\boldsymbol{\mathcal{R}}}_A(:,j,k)^H \Big) \circ \Big(\sum_{j=1}^{r_2} \bar{\boldsymbol{\mathcal{L}}}_B(:,j,k)\bar{\boldsymbol{\mathcal{R}}}_B(:,j,k)^H \Big) \\
&= \sum_{j=1}^{r_1} \sum_{j'=1}^{r_2} \big(\bar{\boldsymbol{\mathcal{L}}}_A(:,j,k) \circ \bar{\boldsymbol{\mathcal{L}}}_B(:,j',k) \big) \cdot \big(\bar{\boldsymbol{\mathcal{R}}}_A(:,j,k) \circ \bar{\boldsymbol{\mathcal{R}}}_B(:,j',k) \big)^H.
\end{aligned}
$$

We can then find an upper bound of $\|(\boldsymbol{\mathcal{L}}_A *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_A^H) \circ (\boldsymbol{\mathcal{L}}_B *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_B^H)\|_*$ via

$$
\begin{aligned}
&\|(\boldsymbol{\mathcal{L}}_A *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_A^H) \circ (\boldsymbol{\mathcal{L}}_B *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_B^H)\|_* \\
&\leq \sum_{k=1}^{n_3} \sum_{j=1}^{r_1} \sum_{j'=1}^{r_2} \frac{1}{\ell} \Big\| \big(\bar{\boldsymbol{\mathcal{L}}}_A(:,j,k) \circ \bar{\boldsymbol{\mathcal{L}}}_B(:,j',k) \big) \cdot \big(\bar{\boldsymbol{\mathcal{R}}}_A(:,j,k) \circ \bar{\boldsymbol{\mathcal{R}}}_B(:,j',k) \big)^H \Big\|_* \\
&= \sum_{k=1}^{n_3} \sum_{j=1}^{r_1} \sum_{j'=1}^{r_2} \frac{1}{\ell} \Big\| \bar{\boldsymbol{\mathcal{L}}}_A(:,j,k) \circ \bar{\boldsymbol{\mathcal{L}}}_B(:,j',k) \Big\|_2 \Big\| \bar{\boldsymbol{\mathcal{R}}}_A(:,j,k) \circ \bar{\boldsymbol{\mathcal{R}}}_B(:,j',k) \Big\|_2 \\
&= \sum_{k=1}^{n_3} \sum_{j=1}^{r_1} \sum_{j'=1}^{r_2} \frac{1}{\ell} \sqrt{\sum_{i=1}^{n_1} |(\bar{\boldsymbol{\mathcal{L}}}_A)_{i,j,k}|^2 |(\bar{\boldsymbol{\mathcal{L}}}_B)_{i,j',k}|^2} \sqrt{\sum_{i=1}^{n_2} |(\bar{\boldsymbol{\mathcal{R}}}_A)_{i,j,k}|^2 |(\bar{\boldsymbol{\mathcal{R}}}_B)_{i,j',k}|^2},
\end{aligned}
$$

where we replace nuclear norm by vector ℓ_2 norms in the second last line because the summands are all rank one matrices. Applying Cauchy-Schwarz inequality twice, we have

$$
\begin{aligned}
&\|(\boldsymbol{\mathcal{L}}_A *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_A^H) \circ (\boldsymbol{\mathcal{L}}_B *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_B^H)\|_* \\
&\leq \sum_{k=1}^{n_3} \frac{1}{\ell} \sqrt{\sum_{j=1}^{r_1} \sum_{j'=1}^{r_2} \sum_{i=1}^{n_1} |(\bar{\boldsymbol{\mathcal{L}}}_A)_{i,j,k}|^2 |(\bar{\boldsymbol{\mathcal{L}}}_B)_{i,j',k}|^2} \sqrt{\sum_{j=1}^{r_1} \sum_{j'=1}^{r_2} \sum_{i=1}^{n_2} |(\bar{\boldsymbol{\mathcal{R}}}_A)_{i,j,k}|^2 |(\bar{\boldsymbol{\mathcal{R}}}_B)_{i,j',k}|^2} \\
&= \sum_{k=1}^{n_3} \frac{1}{\ell} \sqrt{\sum_{i=1}^{n_1} \|\bar{\boldsymbol{\mathcal{L}}}_A(i,:,k)\|_2^2 \|\bar{\boldsymbol{\mathcal{L}}}_B(i,:,k)\|_2^2} \sqrt{\sum_{i=1}^{n_2} \|\bar{\boldsymbol{\mathcal{R}}}_A(i,:,k)\|_2^2 \|\bar{\boldsymbol{\mathcal{R}}}_B(i,:,k)\|_2^2} \\
&\leq \frac{1}{\ell} \sqrt{\sum_{k=1}^{n_3} \sum_{i=1}^{n_1} \|\bar{\boldsymbol{\mathcal{L}}}_A(i,:,k)\|_2^2 \|\bar{\boldsymbol{\mathcal{L}}}_B(i,:,k)\|_2^2} \sqrt{\sum_{k=1}^{n_3} \sum_{i=1}^{n_2} \|\bar{\boldsymbol{\mathcal{R}}}_A(i,:,k)\|_2^2 \|\bar{\boldsymbol{\mathcal{R}}}_B(i,:,k)\|_2^2} \\
&= \ell \sqrt{\sum_{k=1}^{n_3} \sum_{i=1}^{n_1} \|\boldsymbol{\mathcal{L}}_A(i,:,k)\|_2^2 \|\boldsymbol{\mathcal{L}}_B(i,:,k)\|_2^2} \sqrt{\sum_{k=1}^{n_3} \sum_{i=1}^{n_2} \|\boldsymbol{\mathcal{R}}_A(i,:,k)\|_2^2 \|\boldsymbol{\mathcal{R}}_B(i,:,k)\|_2^2} \\
&\leq \ell \Big(\|\boldsymbol{\mathcal{L}}_A\|_{2,2,\infty} \|\boldsymbol{\mathcal{L}}_B\|_F \wedge \|\boldsymbol{\mathcal{L}}_A\|_F \|\boldsymbol{\mathcal{L}}_B\|_{2,2,\infty} \Big) \Big(\|\boldsymbol{\mathcal{R}}_A\|_{2,2,\infty} \|\boldsymbol{\mathcal{R}}_B\|_F \wedge \|\boldsymbol{\mathcal{R}}_A\|_F \|\boldsymbol{\mathcal{R}}_B\|_{2,2,\infty} \Big).
\end{aligned}
\tag{68}
$$

Putting (67) and (68) together, we have

$$|p^{-1}\langle \mathcal{P}_{\Omega}(\mathcal{L}_A *_{\Phi} \mathcal{R}_A^H), \mathcal{P}_{\Omega}(\mathcal{L}_B *_{\Phi} \mathcal{R}_B^H)\rangle - \langle \mathcal{L}_A *_{\Phi} \mathcal{R}_A^H, \mathcal{L}_B *_{\Phi} \mathcal{R}_B^H\rangle|$$

$$\leq c\ell^{\frac{3}{2}}\sqrt{\frac{(n_1 \vee n_2)\log((n_1 \vee n_2)n_3)}{p}}$$

$$\left(\|\mathcal{L}_A\|_{2,2,\infty}\|\mathcal{L}_B\|_F \wedge \|\mathcal{L}_A\|_F\|\mathcal{L}_B\|_{2,2,\infty}\right)\left(\|\mathcal{R}_A\|_{2,2,\infty}\|\mathcal{R}_B\|_F \wedge \|\mathcal{R}_A\|_F\|\mathcal{R}_B\|_{2,2,\infty}\right).$$

D.3 Proof of Lemma 25

In view of Lemma 43, we have

$$\text{dist}(\widetilde{\mathcal{F}}_0, \mathcal{F}_\star) \leq \sqrt{\sqrt{2}+1}\|\mathcal{U}_0 *_{\Phi} \mathcal{G}_0 *_{\Phi} \mathcal{V}_0^H - \mathcal{X}_\star\|_F = \sqrt{\frac{\sqrt{2}+1}{\ell}}\|\bar{U}_0\bar{G}_0\bar{V}_0^H - \bar{X}_\star\|_F$$

$$\leq \sqrt{\frac{(\sqrt{2}+1)2s_r}{\ell}}\|\bar{U}_0\bar{G}_0\bar{V}_0^H - \bar{X}_\star\| = \sqrt{\frac{(\sqrt{2}+1)2s_r}{\ell}}\|\mathcal{U}_0 *_{\Phi} \mathcal{G}_0 *_{\Phi} \mathcal{V}_0^H - \mathcal{X}_\star\|,$$

where we use the fact that $\bar{U}_0\bar{G}_0\bar{V}_0^H - \bar{X}_\star$ has rank at most $2s_r$. Applying the triangle inequality, we obtain

$$\|\mathcal{U}_0 *_{\Phi} \mathcal{G}_0 *_{\Phi} \mathcal{V}_0^H - \mathcal{X}_\star\| \leq \|p^{-1}\mathcal{P}_{\Omega}(\mathcal{X}_\star) - \mathcal{U}_0 *_{\Phi} \mathcal{G}_0 *_{\Phi} \mathcal{V}_0^H\| + \|p^{-1}\mathcal{P}_{\Omega}(\mathcal{X}_\star) - \mathcal{X}_\star\|$$

$$\leq 2\|p^{-1}\mathcal{P}_{\Omega}(\mathcal{X}_\star) - \mathcal{X}_\star\|,$$

where the second inequality relies on the fact that $\mathcal{U}_0 *_{\Phi} \mathcal{G}_0 *_{\Phi} \mathcal{V}_0^H$ is the best tubal rank-r approximation to $p^{-1}\mathcal{P}_{\Omega}(\mathcal{X}_\star)$, i.e., $\|p^{-1}\mathcal{P}_{\Omega}(\mathcal{X}_\star) - \mathcal{U}_0 *_{\Phi} \mathcal{G}_0 *_{\Phi} \mathcal{V}_0^H\| \leq \|p^{-1}\mathcal{P}_{\Omega}(\mathcal{X}_\star) - \mathcal{X}_\star\|$. Combining the above two inequalities yields

$$\text{dist}(\widetilde{\mathcal{F}}_0, \mathcal{F}_\star) \leq 2\sqrt{\frac{(\sqrt{2}+1)2s_r}{\ell}}\|p^{-1}\mathcal{P}_{\Omega}(\mathcal{X}_\star) - \mathcal{X}_\star\| \leq 5\sqrt{\frac{s_r}{\ell}}\|p^{-1}\mathcal{P}_{\Omega}(\mathcal{X}_\star) - \mathcal{X}_\star\|.$$

Using Lemma 53, we know that

$$\|(p^{-1}\mathcal{P}_{\Omega} - \mathcal{I}_{n_1})(\mathcal{X}_\star)\| \leq c\left(\frac{\sqrt{\ell}\log((n_1 \vee n_2)n_3)}{p}\|\mathcal{X}_\star\|_\infty + \sqrt{\frac{\ell\log((n_1 \vee n_2)n_3)}{p}}\|\mathcal{X}_\star\|_{\infty,2}\right),$$

holds with high probability. The proof is finished by applying Lemma 35 and Lemma 36 and plugging the following bounds from incoherence assumption of $\mathcal{X}_\star$:

$$\|\mathcal{X}_\star\|_\infty \leq \sqrt{\ell}\|\mathcal{U}_\star\|_{2,\infty}\|\mathcal{G}_\star\|\|\mathcal{V}_\star\|_{2,\infty} \leq \frac{\mu s_r}{\sqrt{n_1 n_2 \ell n_3}}\kappa\bar{\sigma}_{s_r}(\mathcal{X}_\star);$$

$$\|\mathcal{X}_\star\|_{\infty,2} \leq \|\mathcal{U}_\star\|_{2,\infty}\|\mathcal{G}_\star\|\|\mathcal{V}_\star\| \vee \|\mathcal{U}_\star\|\|\mathcal{G}_\star\|\|\mathcal{V}_\star\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{(n_1 \wedge n_2)n_3 \ell}}\kappa\bar{\sigma}_{s_r}(\mathcal{X}_\star).$$

Appendix E. Proof for Robust Tensor Completion

Lemma 58 *Suppose that $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a fixed tensor and $\boldsymbol{\Omega} \sim \mathrm{Ber}(p)$. Then with high probability,*

$$\|(p^{-1}\boldsymbol{\mathcal{P}_\Omega} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{Z}})\|_{2,\infty} \leq 2\sqrt{\frac{c \log(n_2 n_3)}{p}} \|\boldsymbol{\mathcal{Z}}\|_{2,\infty} + \frac{c \log(n_2 n_3)}{p} \|\boldsymbol{\mathcal{Z}}\|_\infty,$$

for some numerical constant $c > 0$.

Proof For any fixed $b \in [n_1]$, the b-th horizontal slice of the tensor $(p^{-1}\boldsymbol{\mathcal{P}_\Omega} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{Z}})$ can be written as

$$[(p^{-1}\boldsymbol{\mathcal{P}_\Omega} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{Z}})]^H *_{\boldsymbol{\Phi}} \mathring{\mathbf{e}}_b = \sum_{i,j,k}(p^{-1}\delta_{ijk} - 1)\boldsymbol{\mathcal{Z}}_{i,j,k}\bar{\mathbf{e}}_{ijk}^H *_{\boldsymbol{\Phi}} \mathring{\mathbf{e}}_b \coloneqq \sum_{i,j,k}\boldsymbol{\mathcal{H}}_{ijk},$$

where $\boldsymbol{\mathcal{H}}_{ijk}$'s are independent row tensors of size $n_2 \times 1 \times n_3$ and $\mathbb{E}[\bar{\boldsymbol{H}}_{ijk}] = \mathbf{0}$. Let $\boldsymbol{h}_{ijk} \in \mathbb{R}^{n_2 n_3}$ be the column vector obtained by vectorizing $\boldsymbol{\mathcal{H}}_{ijk}$. Then we have

$$\|\boldsymbol{h}_{ijk}\|_2 \leq p^{-1}\boldsymbol{\mathcal{Z}}_{i,j,k}\|\bar{\mathbf{e}}_{ijk}^H *_{\boldsymbol{\Phi}} \mathring{\mathbf{e}}_b\|_F \leq p^{-1}\|\boldsymbol{\mathcal{Z}}\|_\infty.$$

We also have

$$\left|\sum_{i,j,k}\mathbb{E}[\boldsymbol{h}_{ijk}^H\boldsymbol{h}_{ijk}]\right| = \left|\sum_{i,j,k}\mathbb{E}[\|\boldsymbol{\mathcal{H}}_{ijk}\|_F^2]\right| = \frac{1-p}{p}\sum_{i,j,k}\boldsymbol{\mathcal{Z}}_{i,j,k}^2\|\bar{\mathbf{e}}_{ijk}^H *_{\boldsymbol{\Phi}} \mathring{\mathbf{e}}_b\|_F^2.$$

Note that $\bar{\mathbf{e}}_{ijk}^H *_{\boldsymbol{\Phi}} \mathring{\mathbf{e}}_b = \mathbf{0}$ if $i \neq b$. We further have

$$\left|\sum_{i,j,k}\mathbb{E}[\boldsymbol{h}_{ijk}^H\boldsymbol{h}_{ijk}]\right| = \frac{1-p}{p}\sum_{i,j,k}\boldsymbol{\mathcal{Z}}_{i,j,k}^2\|\bar{\mathbf{e}}_{ijk}^H *_{\boldsymbol{\Phi}} \mathring{\mathbf{e}}_b\|_F^2 = \frac{1-p}{p}\sum_{j,k}\boldsymbol{\mathcal{Z}}_{b,j,k}^2\|\bar{\mathbf{e}}_{bjk}^H\|_F^2 \leq p^{-1}\|\boldsymbol{\mathcal{Z}}\|_{2,\infty}^2.$$

We can bound $\|\sum_{i,j,k}\mathbb{E}[\boldsymbol{h}_{ijk}\boldsymbol{h}_{ijk}^H]\|$ by the same quantity in a similar manner. Treating $\boldsymbol{h}_{ijk}$'s as $n_2 n_3 \times 1$ matrices and applying the matrix Bernstein inequality in Lemma 52 gives that with high probability,

$$\|[(p^{-1}\boldsymbol{\mathcal{P}_\Omega} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{Z}})]^H *_{\boldsymbol{\Phi}} \mathring{\mathbf{e}}_b\|_F = \|\boldsymbol{\mathcal{H}}_{ijk}\|_F = \|\boldsymbol{h}_{ijk}\|_F$$

$$\leq 2\sqrt{\frac{c \log(n_2 n_3)}{p}} \|\boldsymbol{\mathcal{Z}}\|_{2,\infty} + \frac{c \log(n_2 n_3)}{p} \|\boldsymbol{\mathcal{Z}}\|_\infty$$

for some numerical constant $c > 0$. ■

E.1 Proof of Lemma 26

We prove Lemma 26 again by induction.

Base case. Since $\rho^0 = 1$, the assumed initial conditions satisfy the base case at $t = 0$.

Induction step. At the t-th iteration, we assume the conditions

$$\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \leq \frac{\epsilon}{\sqrt{\ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star),$$

$$\sqrt{n_1}\|(\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|(\boldsymbol{\mathcal{R}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$$

hold. In view of the condition $\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \leq \frac{0.02}{\sqrt{\ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$ and Lemma 41, one knows that $\boldsymbol{\mathcal{Q}}_t$ exists and $\epsilon := 0.02$.

Distance contraction: By the definition of $\mathrm{dist}^2(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star)$, we have

$$\mathrm{dist}^2(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) \leq \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \|(\boldsymbol{\mathcal{R}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2.$$

According to the update rule (23), we have

$$
\begin{aligned}
&(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
={}& \Big(\boldsymbol{\mathcal{L}}_\sharp - \eta p^{-1} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H + \boldsymbol{\mathcal{S}}_{t+1} - \boldsymbol{\mathcal{X}}_\star - \boldsymbol{\mathcal{S}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} - \boldsymbol{\mathcal{L}}_\star\Big) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
={}& \Big(\boldsymbol{\mathcal{L}}_\triangle - \eta p^{-1} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{\mathcal{X}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} \\
&\qquad - \eta p^{-1} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{S}}_{t+1} - \boldsymbol{\mathcal{S}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1}\Big) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
={}& \boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} - \eta(\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{\mathcal{X}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
&\qquad - \eta(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{\mathcal{X}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
&\qquad - \eta p^{-1} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{S}}_\triangle) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
={}& (1 - \eta)\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} - \eta \boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\triangle^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
&\qquad - \eta(p^{-1}\boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{\mathcal{X}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
&\qquad - \eta p^{-1} \boldsymbol{\mathcal{P}}_{\boldsymbol{\Omega}}(\boldsymbol{\mathcal{S}}_\triangle) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}.
\end{aligned}
\tag{69}
$$

Taking the squared Frobenius norm of both sides of (69) to obtain

$$\|(\mathcal{L}_{t+1} *_{\mathbf{\Phi}} \mathcal{Q}_t - \mathcal{L}_\star) *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2$$

$$=\|(1-\eta)\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}} - \eta\mathcal{L}_\star *_{\mathbf{\Phi}} \mathcal{R}_\triangle^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2$$

$$- 2\eta(1-\eta)\langle\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}, (p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\mathbf{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star) *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\rangle$$

$$+ 2\eta^2\langle\mathcal{L}_\star *_{\mathbf{\Phi}} \mathcal{R}_\triangle^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}},$$

$$(p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\mathbf{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star) *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\rangle$$

$$+ \eta^2\|(p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\mathbf{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star) *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2$$

$$- 2\eta(1-\eta)\langle\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}, p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle) *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\rangle$$

$$+ 2\eta^2\langle\mathcal{L}_\star *_{\mathbf{\Phi}} \mathcal{R}_\triangle^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}},$$

$$p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle) *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\rangle$$

$$+ 2\eta^2\langle(p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\mathbf{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star) *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}},$$

$$p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle) *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\rangle$$

$$+ \eta^2\|p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle) *_{\mathbf{\Phi}} \mathcal{R}_\sharp *_{\mathbf{\Phi}} (\mathcal{R}_\sharp^H *_{\mathbf{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2$$

$$:=\mathfrak{T}_1 - \mathfrak{T}_2 + \mathfrak{T}_3 + \mathfrak{T}_4 - \mathfrak{T}_5 + \mathfrak{T}_6 + \mathfrak{T}_7 + \mathfrak{T}_8.$$

Bound of $\mathfrak{T}_1$, $\mathfrak{T}_2$, $\mathfrak{T}_3$ and $\mathfrak{T}_4$. Repeating the same steps as in the proof of Lemma 24 and utilizing Lemma 48, we have the following:

$$\mathfrak{T}_1 \leq \left((1-\eta)^2 + \frac{2\epsilon}{1-\epsilon}\eta(1-\eta)\right)\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \frac{2\epsilon+\epsilon^2}{(1-\epsilon)^2}\eta^2\|\mathcal{R}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2,$$

$$|\mathfrak{T}_2| \leq 2\eta(1-\eta)\Big(\frac{c_1}{(1-\epsilon)^2}\sqrt{\frac{\mu s_r(n_1+n_2)\log((n_1\vee n_2)n_3)}{pn_1n_2n_3}}\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2$$

$$+ \frac{c_2}{(1-\epsilon)^2}\sqrt{\frac{\ell\log((n_1\vee n_2)n_3)}{p(n_1\wedge n_2)}}\frac{\mu s_r}{n_3}(\rho^t+4)\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\|\mathcal{R}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big),$$

$$|\mathfrak{T}_3| \leq 2\eta^2\Big(\frac{c_1}{(1-\epsilon)^2}\sqrt{\frac{\mu s_r(n_1+n_2)\log((n_1\vee n_2)n_3)}{pn_1n_2n_3}}\|\mathcal{R}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2$$

$$+ \frac{c_2}{(1-\epsilon)^2}\sqrt{\frac{\ell\log((n_1\vee n_2)n_3)}{p(n_1\wedge n_2)}}\frac{2\mu s_r}{n_3}\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\|\mathcal{R}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big),$$

and $\quad \sqrt{\mathfrak{T}_4} \leq \eta\Big(\frac{c_1}{(1-\epsilon)^2}\sqrt{\frac{\mu s_r(n_1+n_2)\log((n_1\vee n_2)n_3)}{pn_1n_2n_3}}\|\mathcal{L}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F$

$$+ \frac{c_2}{(1-\epsilon)^2}\sqrt{\frac{\ell\log((n_1\vee n_2)n_3)}{p(n_1\wedge n_2)}}\frac{\mu s_r}{n_3}(\rho^t+4)\|\mathcal{R}_\triangle *_{\mathbf{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big).$$

Bound of $\mathfrak{T}_5$. Lemma 22 implies $\hat{\mathcal{S}} := \mathcal{P}_{\Omega}(\mathcal{S}_{\triangle})$ is an αp-sparse tensor, hence

$$
\begin{aligned}
&2\eta(1-\eta)\langle \mathcal{L}_{\triangle} *_{\Phi} \mathcal{G}_{\star}^{\frac{1}{2}}, p^{-1}\mathcal{P}_{\Omega}(\mathcal{S}_{\triangle}) *_{\Phi} \mathcal{R}_{\sharp} *_{\Phi} (\mathcal{R}_{\sharp}^H *_{\Phi} \mathcal{R}_{\sharp})^{-1} *_{\Phi} \mathcal{G}_{\star}^{\frac{1}{2}}\rangle \\
=&2\frac{\eta(1-\eta)}{p\ell}\left|\mathrm{tr}\left(\bar{\hat{S}}\bar{R}_{\sharp}(\bar{R}_{\sharp}^H\bar{R}_{\sharp})^{-1}\bar{G}_{\star}\bar{L}_{\triangle}^H\right)\right| \\
\leq&2\frac{\eta(1-\eta)}{p\ell}\|\bar{\hat{S}}\|\|\bar{R}_{\sharp}(\bar{R}_{\sharp}^H\bar{R}_{\sharp})^{-1}\bar{G}_{\star}\bar{L}_{\triangle}^H\|_* \\
\leq&\frac{\alpha\eta(1-\eta)}{\sqrt{\ell}}(n_1+n_2n_3)\|\hat{S}\|_{\infty}\sqrt{s_r}\|\bar{R}_{\sharp}(\bar{R}_{\sharp}^H\bar{R}_{\sharp})^{-1}\bar{G}_{\star}\bar{L}_{\triangle}^H\|_F \\
\leq&\frac{\alpha\eta(1-\eta)}{\sqrt{\ell}}(n_1+n_2n_3)\|\hat{S}\|_{\infty}\sqrt{s_r}\|\bar{R}_{\sharp}(\bar{R}_{\sharp}^H\bar{R}_{\sharp})^{-1}\bar{G}_{\star}^{\frac{1}{2}}\|\|\bar{L}_{\triangle}\bar{G}_{\star}^{\frac{1}{2}}\|_F \\
\leq&2\frac{\alpha\sqrt{s_r}\eta(1-\eta)}{1-\epsilon}(n_1+n_2n_3)\|\mathcal{X}_t-\mathcal{X}_{\star}\|_{\infty}\|\mathcal{L}_{\triangle} *_{\Phi} \mathcal{G}_{\star}^{\frac{1}{2}}\|_F \\
\leq&6\frac{\alpha\mu s_r^{1.5}\eta(1-\eta)}{\ell n_3\sqrt{n_1n_2}}\frac{\epsilon}{1-\epsilon}(n_1+n_2n_3)\rho^{2t}\bar{\sigma}_{s_r}^2(\mathcal{X}_{\star}),
\end{aligned}
$$

where the last step uses the results from Lemma 48 and Lemma 50.

Bound of $\mathfrak{T}_6$. Similar to $\mathfrak{T}_5$, we have

$$
\begin{aligned}
&2\eta^2\langle \mathcal{L}_{\star} *_{\Phi} \mathcal{R}_{\triangle}^H *_{\Phi} \mathcal{R}_{\sharp} *_{\Phi} (\mathcal{R}_{\sharp}^H *_{\Phi} \mathcal{R}_{\sharp})^{-1} *_{\Phi} \mathcal{G}_{\star}^{\frac{1}{2}}, \\
&\qquad\qquad p^{-1}\mathcal{P}_{\Omega}(\mathcal{S}_{\triangle}) *_{\Phi} \mathcal{R}_{\sharp} *_{\Phi} (\mathcal{R}_{\sharp}^H *_{\Phi} \mathcal{R}_{\sharp})^{-1} *_{\Phi} \mathcal{G}_{\star}^{\frac{1}{2}}\rangle \\
=&2\frac{\eta^2}{p\ell}\left|\mathrm{tr}\left(\bar{\hat{S}}\bar{R}_{\sharp}(\bar{R}_{\sharp}^H\bar{R}_{\sharp})^{-1}\bar{G}_{\star}(\bar{R}_{\sharp}^H\bar{R}_{\sharp})^{-1}\bar{R}_{\sharp}^H\bar{R}_{\triangle}\bar{L}_{\star}^H\right)\right| \\
\leq&2\frac{\eta^2}{p\ell}\|\bar{\hat{S}}\|\|\bar{R}_{\sharp}(\bar{R}_{\sharp}^H\bar{R}_{\sharp})^{-1}\bar{G}_{\star}(\bar{R}_{\sharp}^H\bar{R}_{\sharp})^{-1}\bar{R}_{\sharp}^H\bar{R}_{\triangle}\bar{L}_{\star}^H\|_* \\
\leq&\frac{\alpha\eta^2}{\sqrt{\ell}}(n_1+n_2n_3)\|\hat{S}\|_{\infty}\sqrt{s_r}\|\bar{R}_{\sharp}(\bar{R}_{\sharp}^H\bar{R}_{\sharp})^{-1}\bar{G}_{\star}(\bar{R}_{\sharp}^H\bar{R}_{\sharp})^{-1}\bar{R}_{\sharp}^H\bar{R}_{\triangle}\bar{L}_{\star}^H\|_F \\
\leq&\frac{\alpha\eta^2}{\sqrt{\ell}}(n_1+n_2n_3)\|\hat{S}\|_{\infty}\sqrt{s_r}\|\bar{R}_{\sharp}(\bar{R}_{\sharp}^H\bar{R}_{\sharp})^{-1}\bar{G}_{\star}^{\frac{1}{2}}\|^2\|\bar{R}_{\triangle}\bar{G}_{\star}^{\frac{1}{2}}\|_F\|\bar{U}_{\star}\| \\
\leq&2\alpha\eta^2(n_1+n_2n_3)\sqrt{s_r}\frac{1}{(1-\epsilon)^2}\|\mathcal{X}_t-\mathcal{X}_{\star}\|_{\infty}\|\mathcal{R}_{\triangle} *_{\Phi} \mathcal{G}_{\star}^{\frac{1}{2}}\|_F \\
\leq&6\frac{\alpha\mu s_r^{1.5}\eta^2}{\ell n_3\sqrt{n_1n_2}}(n_1+n_2n_3)\frac{\epsilon}{(1-\epsilon)^2}\rho^{2t}\bar{\sigma}_{s_r}^2(\mathcal{X}_{\star}).
\end{aligned}
$$

Bound of $\mathfrak{T}_7$. Using the decomposition $\mathcal{L}_\sharp *_\Phi \mathcal{R}_\sharp^H - \mathcal{X}_\star = \mathcal{L}_\triangle *_\Phi \mathcal{R}_\star^H + \mathcal{L}_\sharp *_\Phi \mathcal{R}_\triangle^H$ and applying the triangle inequality to obtain

$$|\mathfrak{T}_7| = 2\eta^2 \Big| \langle (p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_\Phi \mathcal{R}_\sharp^H - \mathcal{X}_\star) *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}},$$

$$p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle) *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}} \rangle \Big|$$

$$\leq 2\eta^2 \Big(\Big| \langle p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle) *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}},$$

$$(p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\triangle *_\Phi \mathcal{R}_\star^H) *_\Phi \mathcal{R}_\star *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}} \rangle \Big|$$

$$+ \Big| \langle p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle) *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}},$$

$$(p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\triangle *_\Phi \mathcal{R}_\star^H) *_\Phi \mathcal{R}_\triangle *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}} \rangle \Big|$$

$$+ \Big| \langle p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle) *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}},$$

$$(p^{-1}\mathcal{P}_\Omega - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_\Phi \mathcal{R}_\triangle^H) *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}} \rangle \Big| \Big)$$

$$:= 2\eta^2 (\mathfrak{T}_{7,1} + \mathfrak{T}_{7,2} + \mathfrak{T}_{7,3}).$$

For the first term $\mathfrak{T}_{7,1}$, we can invoke Corollary 55 to obtain

$$\mathfrak{T}_{7,1} \leq c_1 \sqrt{\frac{\mu s_r (n_1 + n_2) \log((n_1 \vee n_2)n_3)}{p n_1 n_2 n_3}} \|\mathcal{L}_\triangle *_\Phi \mathcal{R}_\star^H\|_F$$

$$\|p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle) *_\Phi \mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\star^H\|_F$$

$$\leq c_1 \sqrt{\frac{\mu s_r (n_1 + n_2) \log((n_1 \vee n_2)n_3)}{p n_1 n_2 n_3}} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{G}_\star^{-\frac{1}{2}} *_\Phi \mathcal{R}_\star^H\|$$

$$\|p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle)\|_F \|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \|\mathcal{G}_\star^{-\frac{1}{2}} *_\Phi \mathcal{R}_\star^H\|$$

$$= c_1 \sqrt{\frac{\mu s_r (n_1 + n_2) \log((n_1 \vee n_2)n_3)}{p^3 n_1 n_2 n_3}} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \sqrt{\alpha p n_1 n_2 n_3} \|\hat{\mathcal{S}}\|_\infty$$

$$\|\mathcal{R}_\sharp *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|$$

$$\leq \frac{c_1}{(1-\epsilon)^3} \sqrt{\frac{\mu \alpha s_r (n_1 + n_2) \log((n_1 \vee n_2)n_3)}{p^2}} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\hat{\mathcal{S}}\|_\infty$$

$$\leq 2\frac{c_1}{(1-\epsilon)^3} \sqrt{\frac{\mu \alpha s_r (n_1 + n_2) \log((n_1 \vee n_2)n_3)}{p^2}} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{X}_t - \mathcal{X}_\star\|_\infty$$

$$\leq 6\frac{c_1 \mu^{1.5} s_r^{1.5} \epsilon}{p\ell n_3 (1-\epsilon)^3} \sqrt{\frac{\alpha (n_1 + n_2) \log((n_1 \vee n_2)n_3)}{n_1 n_2}} \rho^{2t} \bar{\sigma}_{s_r}^2 (\mathcal{X}_\star).$$

In regard to $\mathfrak{T}_{7,2}$, we can invoke Lemma 56 with $\mathcal{L}_A := \mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}$, $\mathcal{R}_A := \mathcal{R}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}$, $\mathcal{L}_B := p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle)$, $\mathcal{R}_B := \mathcal{R}_\triangle *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\sharp^H$ and use

the consequences in Lemma 57 to obtain

$$
\begin{aligned}
\mathfrak{T}_{7,2} &\leq c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2)\log((n_1 \vee n_2)n_3)}{p}} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle)\|_{2,2,\infty} \\
&\qquad \|\mathcal{R}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,2,\infty} \|\mathcal{R}_\triangle *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\sharp^H\|_F \\
&\leq c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2)\log((n_1 \vee n_2)n_3)}{p^3}} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{P}_\Omega(\mathcal{S}_\triangle)\|_{2,2,\infty} \\
&\qquad \|\mathcal{R}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_F \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\sharp^H\| \\
&\leq c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2)\log((n_1 \vee n_2)n_3)}{p^3}} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \sqrt{\alpha p n_2} \|\hat{\mathcal{S}}\|_\infty \\
&\qquad \|\mathcal{R}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_F \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\sharp^H\| \\
&\leq 2c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2)\log((n_1 \vee n_2)n_3)}{p^3}} \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \sqrt{\alpha p n_2} \|\mathcal{X}_t - \mathcal{X}_\star\|_\infty \\
&\qquad \|\mathcal{R}_\star *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_F \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\sharp^H\| \\
&\leq 6 \frac{c_2 \mu^{1.5} s_r^{1.5}}{p n_3 (1-\epsilon)^3} \sqrt{\frac{\alpha \ell \log((n_1 \vee n_2)n_3)}{(n_1 \wedge n_2)n_3}} \rho^t \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F.
\end{aligned}
$$

Similarly, we can bound $\mathfrak{T}_{7,3}$ as

$$
\begin{aligned}
\mathfrak{T}_{7,3} &\leq c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2)\log((n_1 \vee n_2)n_3)}{p}} \|\mathcal{L}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|p^{-1}\mathcal{P}_\Omega(\mathcal{S}_\triangle)\|_F \\
&\qquad \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\sharp^H\| \\
&\leq c_2 \ell^{\frac{3}{2}} \sqrt{\frac{(n_1 \vee n_2)\log((n_1 \vee n_2)n_3)}{p^3}} \|\mathcal{L}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \sqrt{\alpha p n_1 n_2 n_3} \|\hat{\mathcal{S}}\|_\infty \\
&\qquad \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{R}_\sharp *_\Phi \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty} \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\| \|\mathcal{G}_\star^{\frac{1}{2}} *_\Phi (\mathcal{R}_\sharp^H *_\Phi \mathcal{R}_\sharp)^{-1} *_\Phi \mathcal{R}_\sharp^H\| \\
&\leq 24 \frac{c_2 \mu^2 s_r^2 \epsilon}{p n_3 (1-\epsilon)^3} \sqrt{\frac{\alpha \log((n_1 \vee n_2)n_3)}{\ell (n_1 \wedge n_2)n_3}} \rho^{2t} \bar{\sigma}_{s_r}^2(\mathcal{X}_\star).
\end{aligned}
$$

We then combine the bounds for $\mathfrak{T}_{7,1}$, $\mathfrak{T}_{7,2}$ and $\mathfrak{T}_{7,3}$ to arrive at

$$
\begin{aligned}
|\mathfrak{T}_7| \leq 2\eta^2 \Big(&6 \frac{c_1 \mu^{1.5} s_r^{1.5} \epsilon}{p \ell n_3 (1-\epsilon)^3} \sqrt{\frac{\alpha(n_1 + n_2)\log((n_1 \vee n_2)n_3)}{n_1 n_2}} \rho^{2t} \bar{\sigma}_{s_r}^2(\mathcal{X}_\star) \\
&+ 6 \frac{c_2 \mu^{1.5} s_r^{1.5}}{p n_3 (1-\epsilon)^3} \sqrt{\frac{\alpha \ell \log((n_1 \vee n_2)n_3)}{(n_1 \wedge n_2)n_3}} \rho^t \|\mathcal{L}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \|\mathcal{R}_\triangle *_\Phi \mathcal{G}_\star^{\frac{1}{2}}\|_F \\
&+ 24 \frac{c_2 \mu^2 s_r^2 \epsilon}{p n_3 (1-\epsilon)^3} \sqrt{\frac{\alpha \log((n_1 \vee n_2)n_3)}{\ell (n_1 \wedge n_2)n_3}} \rho^{2t} \bar{\sigma}_{s_r}^2(\mathcal{X}_\star)\Big).
\end{aligned}
$$

73

Bound of $\mathfrak{T}_8$. Similar to $\mathfrak{T}_5$, we have

$$\eta^2 \| p^{-1} \boldsymbol{\mathcal{P}_\Omega}(\boldsymbol{\mathcal{S}_\triangle}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}_\sharp} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}_\sharp^H} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}_\sharp})^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}_\star^{\frac{1}{2}}} \|_F^2$$

$$= \frac{1}{\ell p^2} \eta^2 \| \bar{\hat{\boldsymbol{S}}} \bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}} \|_F^2$$

$$\leq \frac{1}{\ell p^2} \eta^2 s_r \| \bar{\hat{\boldsymbol{S}}} \bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}} \|^2$$

$$\leq \frac{1}{\ell p^2} \eta^2 s_r \| \bar{\hat{\boldsymbol{S}}} \|^2 \| \bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}} \|^2$$

$$\leq \frac{\alpha^2 \eta^2 s_r}{4} (n_1 + n_2 n_3)^2 \| \hat{\boldsymbol{\mathcal{S}}} \|_\infty^2 \| \bar{\boldsymbol{R}}_\sharp (\bar{\boldsymbol{R}}_\sharp^H \bar{\boldsymbol{R}}_\sharp)^{-1} \bar{\boldsymbol{G}}_\star^{\frac{1}{2}} \|^2$$

$$\leq \alpha^2 \eta^2 s_r (n_1 + n_2 n_3)^2 \frac{1}{(1-\epsilon)^2} \| \boldsymbol{\mathcal{X}}_t - \boldsymbol{\mathcal{X}}_\star \|_\infty^2$$

$$\leq 9 \frac{\alpha^2 \mu^2 s_r^3 \eta^2}{\ell n_1 n_2 n_3^2} (n_1 + n_2 n_3)^2 \frac{1}{(1-\epsilon)^2} \rho^{2t} \bar{\sigma}_{s_r}^2 (\boldsymbol{\mathcal{X}}_\star).$$

Taking collectively the bounds for $\mathfrak{T}_1$, $\mathfrak{T}_2$, $\mathfrak{T}_3$, $\mathfrak{T}_4$, $\mathfrak{T}_5$, $\mathfrak{T}_6$, $\mathfrak{T}_7$ and $\mathfrak{T}_8$ together, we have

$$\| (\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \|_F^2$$

$$\leq \left((1-\eta)^2 + \frac{2\epsilon}{1-\epsilon} \eta(1-\eta) \right) \| \boldsymbol{\mathcal{L}_\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \|_F^2 + \frac{2\epsilon + \epsilon^2}{(1-\epsilon)^2} \eta^2 \| \boldsymbol{\mathcal{R}_\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \|_F^2$$

$$+ \eta(1-\eta) \left((2\nu_1 + \nu_3) \| \boldsymbol{\mathcal{L}_\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \|_F^2 + \nu_3 \| \boldsymbol{\mathcal{R}_\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \|_F^2 \right)$$

$$+ \eta^2 \left(\nu_3 \| \boldsymbol{\mathcal{L}_\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \|_F^2 + (2\nu_1 + \nu_3) \| \boldsymbol{\mathcal{R}_\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \|_F^2 \right)$$

$$+ \eta^2 \left(\nu_1(\nu_1 + \nu_3) \| \boldsymbol{\mathcal{L}_\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \|_F^2 + \nu_3(\nu_1 + \nu_3) \| \boldsymbol{\mathcal{R}_\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \|_F^2 \right)$$

$$+ \eta^2 \left(\nu_4 \| \boldsymbol{\mathcal{L}_\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \|_F^2 + \nu_4 \| \boldsymbol{\mathcal{R}_\triangle} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \|_F^2 \right)$$

$$+ 6\eta(1-\eta) \frac{\alpha \mu s_r^{1.5}}{\ell n_3 \sqrt{n_1 n_2}} \frac{\epsilon}{1-\epsilon} (n_1 + n_2 n_3) \rho^{2t} \bar{\sigma}_{s_r}^2 (\boldsymbol{\mathcal{X}}_\star) + 6\eta^2 \frac{\alpha \mu s_r^{1.5}}{\ell n_3 \sqrt{n_1 n_2}} (n_1 + n_2 n_3) \frac{\epsilon}{(1-\epsilon)^2} \rho^{2t} \bar{\sigma}_{s_r}^2 (\boldsymbol{\mathcal{X}}_\star)$$

$$+ 12\eta^2 \frac{c_1 \mu^{1.5} s_r^{1.5} \epsilon}{p \ell n_3 (1-\epsilon)^3} \sqrt{\frac{\alpha(n_1 + n_2) \log((n_1 \vee n_2) n_3)}{n_1 n_2}} \rho^{2t} \bar{\sigma}_{s_r}^2 (\boldsymbol{\mathcal{X}}_\star)$$

$$+ 48\eta^2 \frac{c_2 \mu^2 s_r^2 \epsilon}{p n_3 (1-\epsilon)^3} \sqrt{\frac{\alpha \log((n_1 \vee n_2) n_3)}{\ell(n_1 \wedge n_2) n_3}} \rho^{2t} \bar{\sigma}_{s_r}^2 (\boldsymbol{\mathcal{X}}_\star) + 9\eta^2 \frac{\alpha^2 \mu^2 s_r^3}{\ell n_1 n_2 n_3^2} (n_1 + n_2 n_3)^2 \frac{1}{(1-\epsilon)^2} \rho^{2t} \bar{\sigma}_{s_r}^2 (\boldsymbol{\mathcal{X}}_\star),$$

where ν_1 is defined in (63) and we denote

$$\nu_3 := \frac{c_2}{(1-\epsilon)^2} \sqrt{\frac{\ell \log((n_1 \vee n_2) n_3)}{p(n_1 \wedge n_2)}} \frac{\mu s_r}{n_3} (\rho^t + 4)$$

$$\text{and} \quad \nu_4 := 6 \frac{c_2 \mu^{1.5} s_r^{1.5}}{p n_3 (1-\epsilon)^3} \sqrt{\frac{\alpha \ell \log((n_1 \vee n_2) n_3)}{(n_1 \wedge n_2) n_3}}.$$

A similar bound can be computed for $\|(\mathcal{R}_{t+1} *_{\boldsymbol{\Phi}} \mathcal{Q}_t^{-H} - \mathcal{R}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2$. As a result, we reach the conclusion that

$$\operatorname{dist}^2(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star)$$

$$\leq \hbar^2(\eta; \epsilon, \nu_1, \nu_3, \nu_4)\operatorname{dist}^2(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) + \frac{1}{\ell}\Big(12\eta(1-\eta)\alpha\mu s_r^{1.5}\frac{n_1 + n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{\epsilon}{1-\epsilon}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star)$$

$$+ 12\eta^2\alpha\mu s_r^{1.5}\frac{n_1 + n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{\epsilon}{(1-\epsilon)^2}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star) + 18\eta^2\alpha^2\mu^2 s_r^3\frac{(n_1 + n_2 n_3)^2}{n_1 n_2 n_3^2}\frac{1}{(1-\epsilon)^2}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star)$$

$$+ 24\eta^2\frac{c_1\mu^{1.5}s_r^{1.5}\epsilon}{pn_3(1-\epsilon)^3}\sqrt{\frac{\alpha(n_1 + n_2)\log((n_1 \vee n_2)n_3)}{n_1 n_2}}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star)$$

$$+ 96\eta^2\frac{c_2\mu^2 s_r^2\epsilon}{pn_3(1-\epsilon)^3}\sqrt{\frac{\alpha\ell\log((n_1 \vee n_2)n_3)}{(n_1 \wedge n_2)n_3}}\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star)\Big),$$

where the contraction rate $\hbar^2(\eta; \epsilon, \nu_1, \nu_3, \nu_4)$ is given by

$$\hbar^2(\eta; \epsilon, \nu_1, \nu_3, \nu_4)$$

$$:= (1-\eta)^2 + \Big(\frac{2\epsilon}{1-\epsilon} + 2(\nu_1 + \nu_3)\Big)\eta(1-\eta) + \Big(\frac{2\epsilon + \epsilon^2}{(1-\epsilon)^2} + 2(\nu_1 + \nu_3 + \nu_4) + (\nu_1 + \nu_3)^2\Big)\eta^2.$$

As long as $p \geq c\Big(\frac{\mu^{1.5}s_r^{1.5}(n_1+n_2)\log((n_1 \vee n_2)n_3)}{n_3\sqrt{n_1 n_2}} \vee \frac{\mu^2 s_r^2\ell\log((n_1 \vee n_2)n_3)}{n_3\sqrt{(n_1 \wedge n_2)n_3}}\Big)$ for some sufficiently large constant c and $\alpha\mu s_r^{1.5}\frac{n_1+n_2 n_3}{n_3\sqrt{n_1 n_2}} \leq \alpha\mu s_r^{1.5}\frac{n_1+n_2 n_3}{n_3\sqrt{n_{(2)}n_2}} \leq 10^{-4}$, we have $\nu_1 + \nu_2 + \nu_4 \leq 0.1$ and

$$\frac{c_1\mu^{1.5}s_r^{1.5}}{pn_3}\sqrt{\frac{\alpha(n_1 + n_2)\log((n_1 \vee n_2)n_3)}{n_1 n_2}} + 4\frac{c_2\mu^2 s_r^2}{pn_3}\sqrt{\frac{\alpha\ell\log((n_1 \vee n_2)n_3)}{(n_1 \wedge n_2)n_3}} \leq 0.0005$$

under the setting $\epsilon = 0.02$. With $\frac{1}{5} < \eta \leq \frac{1}{2}$, we can have

$$\operatorname{dist}^2(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star)$$

$$\leq (1 - 0.6\eta)^2\operatorname{dist}^2(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) + \Big(12\eta(1-\eta)\alpha\mu s_r^{1.5}\frac{n_1 + n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{1}{\epsilon(1-\epsilon)} + 0.012\eta^2\frac{1}{\epsilon(1-\epsilon)^3}$$

$$+ 12\eta^2\alpha\mu s_r^{1.5}\frac{n_1 + n_2 n_3}{n_3\sqrt{n_1 n_2}}\frac{1}{\epsilon(1-\epsilon)^2} + 18\eta^2\alpha^2\mu^2 s_r^3\frac{(n_1 + n_2 n_3)^2}{n_1 n_2 n_3^2}\frac{1}{\epsilon^2(1-\epsilon)^2}\Big)\frac{1}{\ell}\epsilon^2\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star)$$

$$\leq (1 - 0.3\eta)^2\frac{1}{\ell}\epsilon^2\rho^{2t}\bar{\sigma}_{s_r}^2(\boldsymbol{\mathcal{X}}_\star).$$

Thus we conclude that

$$\operatorname{dist}(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) \leq \frac{\epsilon}{\sqrt{\ell}}\rho^{t+1}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$$

by setting $\rho = 1 - 0.3\eta$.

Incoherence condition: We first use (69) again to obtain

$$\|(\mathcal{L}_{t+1} *_{\boldsymbol{\Phi}} \mathcal{Q}_t - \mathcal{L}_\star) *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$

$$\leq (1-\eta)\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} + \eta\|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$

$$+ \eta\|(p^{-1}\mathcal{P}_{\boldsymbol{\Omega}} - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$

$$+ \eta\|p^{-1}\mathcal{P}_{\boldsymbol{\Omega}}(\mathcal{S}_\triangle) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$

$$:= \mathfrak{W}_1 + \mathfrak{W}_2 + \mathfrak{W}_3 + \mathfrak{W}_4.$$

Bound of $\mathfrak{W}_1$ and $\mathfrak{W}_2$. The bounds of $\mathfrak{W}_1$ and $\mathfrak{W}_2$ are identical to those of $\mathfrak{B}_1$ and $\mathfrak{B}_2$ in the proof of Lemma 20, respectively, i.e., $\mathfrak{W}_1 \leq (1-\eta)\sqrt{\frac{\mu s_r}{n_1 n_3 \ell}}\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star)$ and $\mathfrak{W}_2 \leq \eta\frac{\epsilon}{1-\epsilon}\sqrt{\frac{\mu s_r}{n_1 n_3 \ell}}\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star)$.

Bound of $\mathfrak{W}_3$. Following Lemma 36 and Lemma 58, we have

$$\mathfrak{W}_3 \leq \eta\|(p^{-1}\mathcal{P}_{\boldsymbol{\Omega}} - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star)\|_{2,\infty}\|\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|$$

$$\leq \frac{\eta}{1-\epsilon}\|(p^{-1}\mathcal{P}_{\boldsymbol{\Omega}} - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star)\|_{2,\infty}$$

$$\leq \frac{\eta}{1-\epsilon}\left(2\sqrt{\frac{c\log(n_2 n_3)}{p}}\|\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star\|_{2,\infty} + \frac{c\log(n_2 n_3)}{p}\|\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star\|_\infty\right)$$

$$= \frac{\eta}{1-\epsilon}\left(2\sqrt{\frac{c\log(n_2 n_3)}{p}}\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H + \mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_{2,\infty} + \frac{c\log(n_2 n_3)}{p}\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H + \mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_\infty\right),$$

where

$$\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H + \mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_{2,\infty}$$

$$\leq \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H\|_{2,\infty} + \|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_{2,\infty}$$

$$\leq \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}\|\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\| + \|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty}\|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|$$

$$\leq \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}(\|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|\|\mathcal{G}_\star^{-1}\| + \|\mathcal{R}_\star *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\|) + \|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{-\frac{1}{2}}\|_{2,\infty}\|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|$$

$$\leq \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}}\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star)(\epsilon\rho^t + 1) + \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}}\epsilon\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star)$$

$$\leq \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}}(2\epsilon + 1)\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star),$$

and $\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H + \mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_\infty \leq 3\frac{\mu s_r}{n_3\sqrt{n_1 n_2 \ell}}\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star)$. Thus

$$\mathfrak{W}_3 \leq \frac{\eta}{1-\epsilon}\left(2\sqrt{\frac{c\log(n_2 n_3)}{p}}\sqrt{\frac{\mu s_r}{n_1 n_3 \ell}}(2\epsilon + 1)\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star) + \frac{c\log(n_2 n_3)}{p}3\frac{\mu s_r}{n_3\sqrt{n_1 n_2 \ell}}\rho^t \bar{\sigma}_{s_r}(\mathcal{X}_\star)\right).$$

76

Bound of $\mathfrak{W}_4$.

$$\mathfrak{W}_4 \leq \eta \|p^{-1}\boldsymbol{\mathcal{P}_{\Omega}}(\boldsymbol{\mathcal{S}}_{\triangle})\|_{2,\infty} \|\boldsymbol{\mathcal{R}}_{\sharp} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_{\sharp}^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_{\sharp})^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\| \leq \frac{\eta}{p} \frac{\sqrt{\alpha p n_2 n_3}}{1-\epsilon} \|\boldsymbol{\mathcal{P}_{\Omega}}(\boldsymbol{\mathcal{S}}_{\triangle})\|_{\infty}$$

$$\leq 6\eta \frac{\sqrt{\alpha \mu s_r}}{1-\epsilon} \sqrt{\frac{\mu s_r}{p n_1 n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_{\star}).$$

Putting all the bounds together, we obtain

$$\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty}$$

$$\leq \left((1-\eta) + \eta \frac{\epsilon}{1-\epsilon} + \frac{\eta}{1-\epsilon} \left(2\sqrt{\frac{c\log(n_2 n_3)}{p}}(2\epsilon + 1) + 3\frac{c\log(n_2 n_3)}{p} \sqrt{\frac{\mu s_r}{n_2 n_3}} + 6\sqrt{\frac{\alpha \mu s_r}{p}} \right) \right) \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_{\star}).$$

We can then have

$$\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_{2,\infty}$$

$$\leq \left((1-\eta) + \eta \frac{\epsilon}{1-\epsilon} + \frac{\eta}{1-\epsilon} \left(2\sqrt{\frac{c\log(n_2 n_3)}{p}}(2\epsilon + 1) + 3\frac{c\log(n_2 n_3)}{p} \sqrt{\frac{\mu s_r}{n_2 n_3}} + 6\sqrt{\frac{\alpha \mu s_r}{p}} \right) \right) \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t.$$

Applying the triangle inequality and Lemma 49, we have

$$\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_{t+1} - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty}$$

$$\leq \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|_{2,\infty} + \left(\|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_{\star}) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_{2,\infty} \right.$$

$$\left. + \|\boldsymbol{\mathcal{L}}_{\star} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{-\frac{1}{2}}\|_{2,\infty} \right) \|\boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-1} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{Q}}_{t+1} - \boldsymbol{\mathcal{Q}}_t) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_{\star}^{\frac{1}{2}}\|$$

$$\leq \left((1-\eta) + \eta \frac{\epsilon}{1-\epsilon} + \frac{\eta}{1-\epsilon} \left(2\sqrt{\frac{c\log(n_2 n_3)}{p}}(2\epsilon + 1) + 3\frac{c\log(n_2 n_3)}{p} \sqrt{\frac{\mu s_r}{n_2 n_3}} \right. \right.$$

$$\left. + 6\sqrt{\frac{\alpha \mu s_r}{p}} \right) \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_{\star}) + \left(\left((1-\eta) + \eta \frac{\epsilon}{1-\epsilon} + \frac{\eta}{1-\epsilon} \left(2\sqrt{\frac{c\log(n_2 n_3)}{p}}(2\epsilon + 1) \right. \right. \right.$$

$$\left. \left. + 3\frac{c\log(n_2 n_3)}{p} \sqrt{\frac{\mu s_r}{n_2 n_3}} + 6\sqrt{\frac{\alpha \mu s_r}{p}} \right) \right) \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t + \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \right) \frac{2\epsilon}{1-\epsilon} \rho^{t+1} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_{\star})$$

$$\leq \left((1-\eta) + \eta \frac{\epsilon}{1-\epsilon} + \frac{\eta}{1-\epsilon} \left(2\sqrt{\frac{c\log(n_2 n_3)}{p}}(2\epsilon + 1) + 3\frac{c\log(n_2 n_3)}{p} \sqrt{\frac{\mu s_r}{n_2 n_3}} + 6\sqrt{\frac{\alpha \mu s_r}{p}} \right) \right.$$

$$+ \frac{2\epsilon}{1-\epsilon} \left((2-\eta) + \eta \frac{\epsilon}{1-\epsilon} + \frac{\eta}{1-\epsilon} \left(2\sqrt{\frac{c\log(n_2 n_3)}{p}}(2\epsilon + 1) + 3\frac{c\log(n_2 n_3)}{p} \sqrt{\frac{\mu s_r}{n_2 n_3}} \right. \right.$$

$$\left. \left. \left. + 6\sqrt{\frac{\alpha \mu s_r}{p}} \right) \right) \right) \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} \rho^t \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_{\star}).$$

Since $p \geq c\frac{\mu^2 s_r^2 \ell \log((n_1 \vee n_2)n_3)}{n_3 \sqrt{(n_1 \wedge n_2)n_3}} \geq c\frac{\mu^2 s_r \ell \log((n_1 \vee n_2)n_3)}{\sqrt{(n_1 \wedge n_2)n_3}}$ for some sufficiently large constant c, we can have $2\sqrt{\frac{c\log(n_2 n_3)}{p}}(2\epsilon + 1) + 3\frac{c\log(n_2 n_3)}{p} \sqrt{\frac{\mu s_r}{n_2 n_3}} \leq 0.1$ with $\epsilon = 0.02$. With $\alpha \mu s_r/p <$

$\alpha\mu s_r^{1.5}\frac{n_1+n_2n_3}{pn_3\sqrt{n_{(2)}n_2}} \le 10^{-4}$, and $\frac{1}{5}\le\eta\le\frac{1}{2}$, we can have the conclusion that $\|(\mathcal{L}_{t+1}*_{\mathbf{\Phi}}\mathcal{Q}_{t+1}-\mathcal{L}_\star)*_{\mathbf{\Phi}}\mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \le (1-0.3\eta)\sqrt{\frac{\mu s_r}{n_1n_3\ell}}\rho^t\bar\sigma_{s_r}(\mathcal{X}_\star)$. We can also obtain a similar bound for $\|(\mathcal{R}_{t+1}*_{\mathbf{\Phi}}\mathcal{Q}_{t+1}^{-H}-\mathcal{R}_\star)*_{\mathbf{\Phi}}\mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$. Thus,

$$\sqrt{n_1}\|(\mathcal{L}_{t+1}*_{\mathbf{\Phi}}\mathcal{Q}_{t+1}-\mathcal{L}_\star)*_{\mathbf{\Phi}}\mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|(\mathcal{R}_{t+1}*_{\mathbf{\Phi}}\mathcal{Q}_{t+1}^{-H}-\mathcal{R}_\star)*_{\mathbf{\Phi}}\mathcal{G}_\star^{\frac{1}{2}}\|_{2,\infty}$$
$$\le \sqrt{\frac{\mu s_r}{n_3\ell}}\rho^{t+1}\bar\sigma_{s_r}(\mathcal{X}_\star)$$

can be achieved by setting $\rho = 1-0.3\eta$. This finishes the proof.

E.2 Proof of Lemma 27

Invoking Lemma 22 with $\mathcal{X}_{-1}=\mathbf{0}$, we have

$$\|\mathcal{P}_{\mathbf{\Omega}}(\mathcal{S}_\star)-\mathcal{S}_0\|_\infty \le 2\frac{\mu s_r}{n_3\sqrt{n_1n_2\ell}}\bar\sigma_1(\mathcal{X}_\star) \quad\text{and}\quad \text{supp}(\mathcal{S}_0)\subseteq\text{supp}(\mathcal{P}_{\mathbf{\Omega}}(\mathcal{S}_\star)),$$

which implies $\mathcal{P}_{\mathbf{\Omega}}(\mathcal{S}_\star)-\mathcal{S}_0$ is an αp-sparse tensor. Let $\mathcal{M}\coloneqq p^{-1}\mathcal{P}_{\mathbf{\Omega}}(\mathcal{Y}-\mathcal{S}_0)$, we have

$$\|\mathcal{X}_\star-\mathcal{M}\| \le \|\mathcal{M}-p^{-1}\mathcal{P}_{\mathbf{\Omega}}(\mathcal{X}_\star)\| + \|p^{-1}\mathcal{P}_{\mathbf{\Omega}}(\mathcal{X}_\star)-\mathcal{X}_\star\|.$$

For the first term, notice that $\mathcal{M}-p^{-1}\mathcal{P}_{\mathbf{\Omega}}(\mathcal{X}_\star)=p^{-1}(\mathcal{P}_{\mathbf{\Omega}}(\mathcal{S}_\star)-\mathcal{S}_0)$. Applying Lemma 34, we have

$$\|\mathcal{M}-p^{-1}\mathcal{P}_{\mathbf{\Omega}}(\mathcal{X}_\star)\| \le \frac{\alpha p\sqrt{\ell}}{2}(n_1+n_2n_3)\|\mathcal{P}_{\mathbf{\Omega}}(\mathcal{S}_\star)-\mathcal{S}_0\|_\infty \le \frac{\alpha p\mu s_r\kappa}{n_3\sqrt{n_1n_2}}(n_1+n_2n_3)\bar\sigma_{s_r}(\mathcal{X}_\star).$$

For the second term, we again have

$$\|p^{-1}\mathcal{P}_{\mathbf{\Omega}}(\mathcal{X}_\star)-\mathcal{X}_\star\| \le c\Big(\frac{\mu s_r\log((n_1\vee n_2)n_3)}{pn_3\sqrt{n_1n_2}} + \sqrt{\frac{\mu s_r\log((n_1\vee n_2)n_3)}{p(n_1\wedge n_2)n_3}}\Big)\kappa\bar\sigma_{s_r}(\mathcal{X}_\star).$$

Since $\mathcal{X}_0=\mathcal{L}_0*_{\mathbf{\Phi}}\mathcal{R}_0^H$ is the best approximation of $\mathcal{M}$ with tubal rank r, we obtain

$$\|\mathcal{X}_\star-\mathcal{X}_0\| \le \|\mathcal{X}_\star-\mathcal{M}\| + \|\mathcal{M}-\mathcal{X}_0\| \le 2\|\mathcal{X}_\star-\mathcal{M}\|$$
$$\le 2\Big(\frac{\alpha p\mu s_r}{n_3\sqrt{n_1n_2}}(n_1+n_2n_3)$$
$$+ c\Big(\frac{\mu s_r\log((n_1\vee n_2)n_3)}{pn_3\sqrt{n_1n_2}} + \sqrt{\frac{\mu s_r\log((n_1\vee n_2)n_3)}{p(n_1\wedge n_2)n_3}}\Big)\Big)\kappa\bar\sigma_{s_r}(\mathcal{X}_\star).$$

Applying Lemma 43 and using the fact that $\bar{X}_\star-\bar{X}_0$ has at most rank-$2s_r$, we obtain

$$\text{dist}(\mathcal{F}_0,\mathcal{F}_\star) \le (\sqrt{2}+1)^{\frac{1}{2}}\frac{1}{\sqrt{\ell}}\|\bar{X}_\star-\bar{X}_0\|_F \le \sqrt{\frac{2(\sqrt{2}+1)s_r}{\ell}}\|\bar{X}_\star-\bar{X}_0\|$$
$$\le 5\sqrt{\frac{s_r}{\ell}}\Big(\frac{\alpha p\mu s_r}{n_3\sqrt{n_1n_2}}(n_1+n_2n_3)$$
$$+ c\Big(\frac{\mu s_r\log((n_1\vee n_2)n_3)}{pn_3\sqrt{n_1n_2}} + \sqrt{\frac{\mu s_r\log((n_1\vee n_2)n_3)}{p(n_1\wedge n_2)n_3}}\Big)\Big)\kappa\bar\sigma_{s_r}(\mathcal{X}_\star).$$

78

As long as $p \geq c\mu s_r \kappa^2 \log((n_1 \vee n_2)n_3)/\sqrt{(n_1 \wedge n_2)n_3}$ for some sufficiently large constant c, we can have $5\sqrt{s_r}\kappa c\left(\frac{\mu s_r \log((n_1 \vee n_2)n_3)}{pn_3\sqrt{n_1 n_2}} + \sqrt{\frac{\mu s_r \log((n_1 \vee n_2)n_3)}{p(n_1 \wedge n_2)n_3}}\right) \leq 0.01$. Given $\epsilon = 5c_0 + 0.01$ and $\alpha/p \leq \frac{c_0}{\mu s_r^{1.5}\kappa \frac{n_1 + n_2 n_3}{n_3\sqrt{n_1 n_2}}}$, our first claim

$$\text{dist}(\boldsymbol{\mathcal{F}}_0, \boldsymbol{\mathcal{F}}_\star) \leq \frac{5c_0 + 0.01}{\sqrt{\ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) \tag{70}$$

is proved.

To prove the second claim, we again define $\boldsymbol{\mathcal{L}}_\sharp := \boldsymbol{\mathcal{L}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_0$, $\boldsymbol{\mathcal{R}}_\sharp := \boldsymbol{\mathcal{R}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_0^{-H}$, $\boldsymbol{\mathcal{L}}_\triangle := \boldsymbol{\mathcal{L}}_\sharp - \boldsymbol{\mathcal{L}}_\star$, $\boldsymbol{\mathcal{R}}_\triangle := \boldsymbol{\mathcal{R}}_\sharp - \boldsymbol{\mathcal{R}}_\star$, $\boldsymbol{\mathcal{S}}_\triangle := \boldsymbol{\mathcal{S}}_0 - \boldsymbol{\mathcal{S}}_\star$. Notice that $\boldsymbol{\mathcal{U}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}_0^H = \text{t-SVD}_r(p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{S}}_0)) = \text{t-SVD}_r(p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}_\star - \boldsymbol{\mathcal{S}}_\triangle))$, thus

$$\boldsymbol{\mathcal{L}}_0 = \boldsymbol{\mathcal{U}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0^{\frac{1}{2}} = p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}_\star - \boldsymbol{\mathcal{S}}_\triangle) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{V}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0^{-\frac{1}{2}} = p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}_\star - \boldsymbol{\mathcal{S}}_\triangle) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_0^{-1}$$
$$= p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}_\star - \boldsymbol{\mathcal{S}}_\triangle) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_0 *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_0^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_0)^{-1}.$$

Multiplying $\boldsymbol{\mathcal{Q}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}$ on both sides, we have

$$\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} = \boldsymbol{\mathcal{L}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} = p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}_\star - \boldsymbol{\mathcal{S}}_\triangle) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_0 *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_0^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_0)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_0 *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}$$
$$= p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}_\star - \boldsymbol{\mathcal{S}}_\triangle) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}.$$

Subtracting $\boldsymbol{\mathcal{X}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}$ on both sides, we have

$$\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} + \boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\triangle^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}$$
$$= (p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}_\star) - \boldsymbol{\mathcal{X}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} - p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{S}}_\triangle) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}},$$

where we use the fact $\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} = \boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}$. Thus,

$$\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \|\boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\triangle^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty}$$
$$+ \|(p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}_\star) - \boldsymbol{\mathcal{X}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty}$$
$$+ \|p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{S}}_\triangle) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty}$$
$$:= \mathfrak{G}_1 + \mathfrak{G}_2 + \mathfrak{G}_3.$$

Bound of $\mathfrak{G}_1$. This term has been controlled as in (56) by $\mathfrak{G}_1 \leq \sqrt{\frac{\mu s_r}{n_1 n_3 \ell}}\frac{\epsilon}{1-\epsilon}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$.

Bound of $\mathfrak{G}_2$. By Lemma 36 and Lemma 58, we have

$$\mathfrak{G}_2 \leq \|(p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{X}}_\star) - \boldsymbol{\mathcal{X}}_\star)\|_{2,\infty}\|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|$$
$$\leq \frac{1}{1-\epsilon}\left(2\sqrt{\frac{c\log(n_2 n_3)}{p}}\|\boldsymbol{\mathcal{X}}_\star\|_{2,\infty} + \frac{c\log(n_2 n_3)}{p}\|\boldsymbol{\mathcal{X}}_\star\|_\infty\right)$$
$$\leq \frac{\kappa}{1-\epsilon}\left(2\sqrt{\frac{c\log(n_2 n_3)}{p}}\sqrt{\frac{\mu s_r}{n_1 n_3 \ell}} + \frac{c\log(n_2 n_3)}{p}\frac{\mu s_r}{\sqrt{n_1 n_2 \ell n_3}}\right)\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$$

holds with high probability.

Bound of $\mathfrak{G}_3$. By Lemmas 34, 36 and 37, we have

$$
\begin{aligned}
\mathfrak{G}_3 &\leq \sqrt{n_2\ell}\|p^{-1}\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{S}}_\triangle)\|_{2,\infty}\|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \\
&\leq p^{-1}\sqrt{n_2\ell}\|\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{S}}_\triangle)\|_{2,\infty}\|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|_{2,\infty}\|\boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\| \\
&\leq p^{-1}\sqrt{n_2\ell}\sqrt{\alpha p n_2 n_3}\|\boldsymbol{\mathcal{P}}_\Omega(\boldsymbol{\mathcal{S}}_\triangle)\|_\infty\|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|_{2,\infty}\|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|^2 \\
&\leq 2n_2\sqrt{\alpha n_3\ell}\frac{\mu s_r}{n_3\sqrt{pn_1n_2\ell}}\bar{\sigma}_1(\boldsymbol{\mathcal{X}}_\star)\|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|_{2,\infty}\|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|^2 \\
&\leq 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha n_2}{pn_1n_3}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)\|\boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|_{2,\infty} \\
&\leq 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha n_2}{pn_1n_3}}\Big(\sqrt{\frac{\mu s_r}{n_2n_3\ell}} + \|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|_{2,\infty}\Big)\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star).
\end{aligned}
$$

Note that $\|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\|_{2,\infty} \leq \frac{\|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty}}{\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)}$. Thus,

$$
\begin{aligned}
\sqrt{n_1}\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{n_3\ell}}\Big(&\frac{\epsilon}{1-\epsilon} + \frac{2\kappa}{1-\epsilon}\sqrt{\frac{c\log(n_2n_3)}{p}} + \frac{c\kappa\log(n_2n_3)}{p(1-\epsilon)}\sqrt{\frac{\mu s_r}{n_2n_3}} \\
&+ 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{pn_3}}\Big)\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) + 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{pn_3}}\sqrt{n_2}\|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty},
\end{aligned}
$$

and similarly one can see

$$
\begin{aligned}
\sqrt{n_2}\|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \leq \sqrt{\frac{\mu s_r}{n_3\ell}}\Big(&\frac{\epsilon}{1-\epsilon} + \frac{2\kappa}{1-\epsilon}\sqrt{\frac{c\log(n_1n_3)}{p}} + \frac{c\kappa\log(n_1n_3)}{p(1-\epsilon)}\sqrt{\frac{\mu s_r}{n_1n_3}} \\
&+ 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{pn_3}}\Big)\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) + 2\frac{\mu s_r \kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{pn_3}}\sqrt{n_1}\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty}.
\end{aligned}
$$

Since $p \geq c\mu s_r \kappa^2 \log((n_1 \vee n_2)n_3)/\sqrt{(n_1 \wedge n_2)n_3}$ for some sufficiently large constant c, we can have $2\kappa\sqrt{\frac{c\log((n_1\vee n_2)n_3)}{p}} + \frac{c\kappa\log((n_1\vee n_2)n_3)}{p}\sqrt{\frac{\mu s_r}{(n_1\wedge n_2)n_3}} \leq 0.1$. Therefore,

$$
\begin{aligned}
&\sqrt{n_1}\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \vee \sqrt{n_2}\|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_{2,\infty} \\
&\leq \frac{\frac{\epsilon}{1-\epsilon} + \frac{2\kappa}{1-\epsilon}\sqrt{\frac{c\log((n_1\vee n_2)n_3)}{p}} + \frac{c\kappa\log((n_1\vee n_2)n_3)}{p(1-\epsilon)}\sqrt{\frac{\mu s_r}{(n_1\wedge n_2)n_3}} + 2\frac{\mu s_r\kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{pn_3}}}{1 - 2\frac{\mu s_r\kappa}{(1-\epsilon)^2}\sqrt{\frac{\alpha}{pn_3}}}\sqrt{\frac{\mu s_r}{n_3\ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) \\
&\leq \frac{\frac{\epsilon}{1-\epsilon} + \frac{0.1}{1-\epsilon} + 2\frac{c_0}{(1-\epsilon)^2}}{1 - 2\frac{c_0}{(1-\epsilon)^2}}\sqrt{\frac{\mu s_r}{n_3\ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) \\
&= \frac{\frac{5c_0}{1-5c_0} + \frac{0.1}{1-5c_0} + 2\frac{c_0}{(1-5c_0)^2}}{1 - 2\frac{c_0}{(1-5c_0)^2}}\sqrt{\frac{\mu s_r}{n_3\ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star) \\
&\leq \sqrt{\frac{\mu s_r}{n_3\ell}}\bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)
\end{aligned}
$$

as long as $c_0 \leq 0.05$. This finishes the proof.

Appendix F. Proof for Tensor Regression

We begin with a useful lemma regarding TRIP, which can be proved in the same way as in Lemma E.7 of Han et al. (2022).

Lemma 59 *Suppose that $\mathcal{A}(\cdot)$ obeys the $2r$-TRIP with a constant δ_{2r}. Then for any $\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of tubal rank at most r, we have*

$$|\langle \mathcal{A}(\boldsymbol{\mathcal{X}}_1), \mathcal{A}(\boldsymbol{\mathcal{X}}_2) \rangle - \langle \boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2 \rangle| \le \delta_{2r} \|\boldsymbol{\mathcal{X}}_1\|_F \|\boldsymbol{\mathcal{X}}_2\|_F,$$

or equivalently,

$$|\langle (\mathcal{A}^*\mathcal{A} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{X}}_1), \boldsymbol{\mathcal{X}}_2 \rangle| \le \delta_{2r} \|\boldsymbol{\mathcal{X}}_1\|_F \|\boldsymbol{\mathcal{X}}_2\|_F. \tag{71}$$

Then following from the variational representation of the Frobenius norm, for any tensor $\boldsymbol{\mathcal{R}} \in \mathbb{R}^{n_2 \times r \times n_3}$, we have

$$\begin{aligned}
\|(\mathcal{A}^*\mathcal{A} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{X}}_1) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}\|_F &= \max_{\widetilde{\boldsymbol{\mathcal{L}}} : \|\widetilde{\boldsymbol{\mathcal{L}}}\|_F \le 1} \langle (\mathcal{A}^*\mathcal{A} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{X}}_1) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}, \widetilde{\boldsymbol{\mathcal{L}}} \rangle \\
&\le \max_{\widetilde{\boldsymbol{\mathcal{L}}} : \|\widetilde{\boldsymbol{\mathcal{L}}}\|_F \le 1} \delta_{2r} \|\boldsymbol{\mathcal{X}}_1\|_F \|\widetilde{\boldsymbol{\mathcal{L}}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}^H\|_F \\
&\le \delta_{2r} \|\boldsymbol{\mathcal{X}}_1\|_F \|\boldsymbol{\mathcal{R}}\|, \tag{72}
\end{aligned}$$

where the last line follows from the relation $\|\boldsymbol{\mathcal{A}} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{B}}\|_F \le \|\boldsymbol{\mathcal{A}}\|_F \|\boldsymbol{\mathcal{B}}\|$.

F.1 Proof of Lemma 28

Since $\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star) \le \frac{0.1}{\sqrt{\ell}} \bar{\sigma}_{s_r}(\boldsymbol{\mathcal{X}}_\star)$, Lemma 41 ensures that $\boldsymbol{\mathcal{Q}}_t$, the optimal alignment tensor between $\boldsymbol{\mathcal{F}}_t$ and $\boldsymbol{\mathcal{F}}_\star$ exists, and $\epsilon := 0.1$. Again, repeating the derivation for Lemma 41, we have

$$\|\boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\| \vee \|\boldsymbol{\mathcal{R}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{-\frac{1}{2}}\| \le \epsilon. \tag{73}$$

The conclusion $\|\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H - \boldsymbol{\mathcal{X}}_\star\|_F \le 1.5\mathrm{dist}(\boldsymbol{\mathcal{F}}_t, \boldsymbol{\mathcal{F}}_\star)$ is a simple consequence of Lemma 45; see (51) for details. In the following, we focus on proving the distance contraction.

First of all, we have by the definition of $\mathrm{dist}(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star)$ that

$$\mathrm{dist}^2(\boldsymbol{\mathcal{F}}_{t+1}, \boldsymbol{\mathcal{F}}_\star) \le \|(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2 + \|(\boldsymbol{\mathcal{R}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t^{-H} - \boldsymbol{\mathcal{R}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}\|_F^2.$$

According to the update rule (28) and the decomposition $\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{\mathcal{X}}_\star = \boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H + \boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\triangle^H$, we have

$$\begin{aligned}
&(\boldsymbol{\mathcal{L}}_{t+1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{Q}}_t - \boldsymbol{\mathcal{L}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
&= \Big(\boldsymbol{\mathcal{L}}_\sharp - \eta \mathcal{A}^*\mathcal{A}(\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{\mathcal{X}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} - \boldsymbol{\mathcal{L}}_\star \Big) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
&= \Big(\boldsymbol{\mathcal{L}}_\triangle - \eta(\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{\mathcal{X}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} \\
&\qquad - \eta(\mathcal{A}^*\mathcal{A} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{\mathcal{X}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} \Big) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
&= (1 - \eta) \boldsymbol{\mathcal{L}}_\triangle *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} - \eta \boldsymbol{\mathcal{L}}_\star *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\triangle^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}} \\
&\qquad - \eta(\mathcal{A}^*\mathcal{A} - \boldsymbol{\mathcal{I}}_{n_1})(\boldsymbol{\mathcal{L}}_\sharp *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp^H - \boldsymbol{\mathcal{X}}_\star) *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp *_{\boldsymbol{\Phi}} (\boldsymbol{\mathcal{R}}_\sharp^H *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{G}}_\star^{\frac{1}{2}}. \tag{74}
\end{aligned}$$

Taking the squared Frobenius norm of both sides of (74) to obtain

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

Bound of $\mathfrak{S}_1$. The first term $\mathfrak{S}_1$ has already been controlled in (53) as follows.

$$
\mathfrak{S}_1 \leq \left((1-\eta)^2 + \frac{2\epsilon}{1-\epsilon}\eta(1-\eta) \right) \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \frac{2\epsilon + \epsilon^2}{(1-\epsilon)^2}\eta^2 \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2.
$$

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

$$
\begin{aligned}
|\mathfrak{S}_2| =&\ 2\eta(1-\eta)\left|\langle \mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}, (\mathcal{A}^*\mathcal{A} - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\rangle\right| \\
\leq&\ 2\eta(1-\eta)\Big(\left|\langle \mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}, (\mathcal{A}^*\mathcal{A} - \mathcal{I}_{n_1})(\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\rangle\right| \\
&+ \left|\langle \mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}, (\mathcal{A}^*\mathcal{A} - \mathcal{I}_{n_1})(\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\rangle\right| \\
&+ \left|\langle \mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}, (\mathcal{A}^*\mathcal{A} - \mathcal{I}_{n_1})(\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\rangle\right| \Big).
\end{aligned}
$$

Applying Lemma 59 to further obtain

$$
\begin{aligned}
|\mathfrak{S}_2| \leq&\ 2\eta(1-\eta)\delta_{2r}\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H\|_F + \|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F + \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F \Big) \\
&\qquad \|\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star *_{\boldsymbol{\Phi}} \mathcal{L}_\triangle^H\|_F \\
\leq&\ 2\eta(1-\eta)\delta_{2r}\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H\|_F + \|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F + \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F \Big) \\
&\qquad \|\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F.
\end{aligned}
$$

Taking the condition (73) and Lemma 44 and Lemma 45 together to obtain

$$
\|\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \leq \frac{1}{1-\epsilon};
$$

$$
\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H\|_F + \|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F + \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F \leq (1 + \frac{\epsilon}{2})\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F + \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \Big).
$$

Hence, we have

$$
\begin{aligned}
|\mathfrak{S}_2| &\leq \eta(1-\eta)\frac{\delta_{2r}(2+\epsilon)}{1-\epsilon}\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F + \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big)\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \\
&= \eta(1-\eta)\frac{\delta_{2r}(2+\epsilon)}{1-\epsilon}\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big) \\
&\leq \eta(1-\eta)\frac{\delta_{2r}(2+\epsilon)}{1-\epsilon}\Big(\frac{3}{2}\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \frac{1}{2}\|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big),
\end{aligned}
$$

where we use the elementary inequality $2ab \leq a^2 + b^2$ in the last line.

Bound of $\mathfrak{S}_3$. The third term $\mathfrak{S}_3$ can be similarly bounded as

$$
\begin{aligned}
|\mathfrak{S}_3| &\leq 2\eta^2\delta_{2r}\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H\|_F + \|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F + \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F\Big) \\
&\qquad \|\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{L}_\star^H\|_F \\
&\leq 2\eta^2\delta_{2r}\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H\|_F + \|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F + \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F\Big) \\
&\qquad \|\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|^2\|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{L}_\star^H\|_F \\
&\leq \eta^2\frac{\delta_{2r}(2+\epsilon)}{(1-\epsilon)^2}\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F + \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big)\|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \\
&\leq \eta^2\frac{\delta_{2r}(2+\epsilon)}{(1-\epsilon)^2}\Big(\frac{1}{2}\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \frac{3}{2}\|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big).
\end{aligned}
$$

Bound of $\mathfrak{S}_4$. For the last term $\mathfrak{S}_4$, we have

$$
\begin{aligned}
\sqrt{\mathfrak{S}_4} &= \eta\|(\mathcal{A}^*\mathcal{A} - \mathcal{I}_{n_1})(\mathcal{L}_\sharp *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp^H - \mathcal{X}_\star) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \\
&\leq \eta\Big(\|(\mathcal{A}^*\mathcal{A} - \mathcal{I}_{n_1})(\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \\
&\qquad + \|(\mathcal{A}^*\mathcal{A} - \mathcal{I}_{n_1})(\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F \\
&\qquad + \|(\mathcal{A}^*\mathcal{A} - \mathcal{I}_{n_1})(\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H) *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big).
\end{aligned}
$$

Following (72), we obtain

$$
\begin{aligned}
\sqrt{\mathfrak{S}_4} &\leq \eta\delta_{2r}\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\star^H\|_F + \|\mathcal{L}_\star *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F + \|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{R}_\triangle^H\|_F\Big)\|\mathcal{R}_\sharp *_{\boldsymbol{\Phi}} (\mathcal{R}_\sharp^H *_{\boldsymbol{\Phi}} \mathcal{R}_\sharp)^{-1} *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\| \\
&\leq \eta\frac{\delta_{2r}(2+\epsilon)}{2(1-\epsilon)}\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F + \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big).
\end{aligned}
$$

We can then take the squares of both sides and use $(a+b)^2 \leq 2a^2 + 2b^2$ to reach

$$
\mathfrak{S}_4 \leq \eta^2\frac{\delta_{2r}^2(2+\epsilon)^2}{2(1-\epsilon)^2}\Big(\|\mathcal{L}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \|\mathcal{R}_\triangle *_{\boldsymbol{\Phi}} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2\Big).
$$

83

Taking the bounds for $\mathfrak{S}_1, \mathfrak{S}_2, \mathfrak{S}_3, \mathfrak{S}_4$ collectively yields

$$\|(\mathcal{L}_{t+1} *_{\Phi} \mathcal{Q}_t - \mathcal{L}_\star) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2$$

$$\leq \Big((1-\eta)^2 + \frac{2\epsilon}{1-\epsilon}\eta(1-\eta)\Big)\|\mathcal{L}_{\triangle} *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \frac{2\epsilon + \epsilon^2}{(1-\epsilon)^2}\eta^2\|\mathcal{R}_{\triangle} *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2$$

$$+ \eta(1-\eta)\frac{\delta_{2r}(2+\epsilon)}{1-\epsilon}\Big(\frac{3}{2}\|\mathcal{L}_{\triangle} *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \frac{1}{2}\|\mathcal{R}_{\triangle} *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big)$$

$$+ \eta^2\frac{\delta_{2r}(2+\epsilon)}{(1-\epsilon)^2}\Big(\frac{1}{2}\|\mathcal{L}_{\triangle} *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \frac{3}{2}\|\mathcal{R}_{\triangle} *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F\Big)$$

$$+ \eta^2\frac{\delta_{2r}^2(2+\epsilon)^2}{2(1-\epsilon)^2}\Big(\|\mathcal{L}_{\triangle} *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \|\mathcal{R}_{\triangle} *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2\Big).$$

A similar bound holds for the second term in (74). Therefore we obtain

$$\|(\mathcal{L}_{t+1} *_{\Phi} \mathcal{Q}_t - \mathcal{L}_\star) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \|(\mathcal{R}_{t+1} *_{\Phi} \mathcal{Q}_t^{-H} - \mathcal{R}_\star) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 \leq \hbar^2(\eta; \epsilon, \delta_{2r})\mathrm{dist}^2(\mathcal{F}_t, \mathcal{F}_\star),$$

where the contraction rate is given by

$$\hbar^2(\eta; \epsilon, \delta_{2r}) := (1-\eta)^2 + \frac{2\epsilon + \delta_{2r}(4+2\epsilon)}{1-\epsilon}\eta(1-\eta) + \frac{2\epsilon + \epsilon^2 + \delta_{2r}(4+2\epsilon) + \delta_{2r}^2(2+\epsilon)^2}{(1-\epsilon)^2}\eta^2.$$

With $\epsilon = 0.1$, $\delta_{2r} \leq 0.02$, and $0 < \eta \leq \frac{2}{3}$, we have $\hbar(\eta; \epsilon, \delta_{2r}) \leq 1 - 0.6\eta$. Thus we conclude that

$$\mathrm{dist}(\mathcal{F}_{t+1}, \mathcal{F}_\star) \leq \sqrt{\|(\mathcal{L}_{t+1} *_{\Phi} \mathcal{Q}_t - \mathcal{L}_\star) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2 + \|(\mathcal{R}_{t+1} *_{\Phi} \mathcal{Q}_t^{-H} - \mathcal{R}_\star) *_{\Phi} \mathcal{G}_\star^{\frac{1}{2}}\|_F^2}$$

$$\leq (1 - 0.6\eta)\mathrm{dist}(\mathcal{F}_t, \mathcal{F}_\star).$$

F.2 Proof of Lemma 29

We first invoke Lemma 43 and use the fact that $\mathcal{L}_0 *_{\Phi} \mathcal{R}_0^H - \mathcal{X}_\star$ has tubal rank at most $2r$ to obtain

$$\mathrm{dist}(\mathcal{F}_0, \mathcal{F}_\star) \leq \sqrt{\sqrt{2}+1}\|\mathcal{L}_0 *_{\Phi} \mathcal{R}_0^H - \mathcal{X}_\star\|_F \leq \sqrt{2(\sqrt{2}+1)}\|\mathcal{L}_0 *_{\Phi} \mathcal{R}_0^H - \mathcal{X}_\star\|_{F,r}.$$

On the other hand, combining Lemma 46 and (72) to reach at

$$\|(\mathcal{A}^*\mathcal{A} - \mathcal{I}_{n_1})(\mathcal{X}_\star)\|_{F,r} = \max_{\widetilde{\mathcal{R}} \in \mathbb{R}^{n_2 \times r \times n_3}: \|\widetilde{\mathcal{R}}\| \leq 1} \|(\mathcal{A}^*\mathcal{A} - \mathcal{I}_{n_1})(\mathcal{X}_\star) *_{\Phi} \widetilde{\mathcal{R}}\|_F \leq \delta_{2r}\|\mathcal{X}_\star\|_F.$$

Note that $\mathcal{L}_0 *_{\Phi} \mathcal{R}_0^H$ is the best tubal rank-r approximation of $\mathcal{A}^*\mathcal{A}(\mathcal{X}_\star)$, we can apply the triangle inequality combined with Lemma 47 to obtain

$$\|\mathcal{L}_0 *_{\Phi} \mathcal{R}_0^H - \mathcal{X}_\star\|_{F,r} \leq \|\mathcal{A}^*\mathcal{A}(\mathcal{X}_\star) - \mathcal{L}_0 *_{\Phi} \mathcal{R}_0^H\|_{F,r} + \|\mathcal{A}^*\mathcal{A}(\mathcal{X}_\star) - \mathcal{X}_\star\|_{F,r}$$

$$\leq 2\|(\mathcal{A}^*\mathcal{A} - \mathcal{I}_{n_1})(\mathcal{X}_\star)\|_{F,r} \leq 2\delta_{2r}\|\mathcal{X}_\star\|_F.$$

As a result, we have

$$\mathrm{dist}(\mathcal{F}_0, \mathcal{F}_\star) \leq 2\sqrt{2(\sqrt{2}+1)}\delta_{2r}\|\mathcal{X}_\star\|_F \leq 5\delta_{2r}\sqrt{\frac{s_r}{\ell}}\kappa\bar{\sigma}_{s_r}(\mathcal{X}_\star).$$

Appendix G. Computational Complexity of Algorithm 1

We provide the breakdown of ScaledGD's computational complexity for tensor RPCA in Algorithm 1:

1. Compute $\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H$ ($\boldsymbol{\mathcal{L}}_t \in \mathbb{R}^{n_1 \times r \times n_3}$ and $\boldsymbol{\mathcal{R}}_t \in \mathbb{R}^{n_2 \times r \times n_3}$): $n_1 n_2 n_3 r + n_1 n_2 n_3^2$ flops.

2. Compute $\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H$: $n_1 n_2 n_3$ flops.

3. Soft-thresholding on $\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H$: $n_1 n_2 n_3$ flops.

4. Compute $\boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H + \boldsymbol{\mathcal{S}}_{t+1} - \boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{S}}_{t+1} - (\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{L}}_t *_{\boldsymbol{\Phi}} \boldsymbol{\mathcal{R}}_t^H)$ and transform the result into the spectral domain: $n_1 n_2 n_3 + n_1 n_2 n_3^2$ flops.

5. Compute $\overline{\boldsymbol{R}}_t^{(k)^H} \overline{\boldsymbol{R}}_t^{(k)}$ ($\overline{\boldsymbol{R}}_t^{(k)} \in \mathbb{C}^{n_2 \times r}$): $n_2 r^2$ flops.

6. Compute $(\overline{\boldsymbol{R}}_t^{(k)^H} \overline{\boldsymbol{R}}_t^{(k)})^{-1}$: $\mathcal{O}(r^3)$ flops.

7. Compute $\overline{\boldsymbol{R}}_t^{(k)} (\overline{\boldsymbol{R}}_t^{(k)^H} \overline{\boldsymbol{R}}_t^{(k)})^{-1}$: $n_2 r^2$ flops.

8. Compute $(\overline{\boldsymbol{L}}_t^{(k)} \overline{\boldsymbol{R}}_t^{(k)^H} + \overline{\boldsymbol{S}}_{t+1}^{(k)} - \overline{\boldsymbol{Y}}^{(k)}) \cdot \overline{\boldsymbol{R}}_t^{(k)} (\overline{\boldsymbol{R}}_t^{(k)^H} \overline{\boldsymbol{R}}_t^{(k)})^{-1}$: $n_1 n_2 r$ flops.

9. Compute $\overline{\boldsymbol{L}}_{t+1}^{(k)} = \overline{\boldsymbol{L}}_t^{(k)} - \eta (\overline{\boldsymbol{L}}^{(k)} \overline{\boldsymbol{R}}^{(k)^H} + \overline{\boldsymbol{S}}^{(k)} - \overline{\boldsymbol{Y}}^{(k)}) \cdot \overline{\boldsymbol{R}}^{(k)} (\overline{\boldsymbol{R}}^{(k)^H} \overline{\boldsymbol{R}}^{(k)})^{-1}$: $2n_1 r$ flops.

10. Repeat step 5 - 9 for computing $\overline{\boldsymbol{R}}_{t+1}^{(k)}$: $2n_1 r^2 + \mathcal{O}(r^3) + n_1 n_2 r + 2n_2 r$ flops.

In total, ScaledGD costs $\mathcal{O}(n_1 n_2 n_3 r + n_1 n_2 n_3^2 + n_1 n_2 n_3 + (n_1 + n_2) n_3 r^2 + n_3 r^3 + (n_1 + n_2) n_3 r) = \mathcal{O}(n_1 n_2 n_3 r + n_1 n_2 n_3^2 + (n_1 + n_2) n_3 r^2 + n_3 r^3)$ flops per iteration provided that $r \ll n_1 \wedge n_2$. We conclude by noting that for some special invertible linear transforms L, e.g., DFT, since the application of DFT on an n_3-dimensional vector requires $\mathcal{O}(n_3 \log(n_3))$ operations, the per-iteration complexity is $\mathcal{O}(n_1 n_2 n_3 r + n_1 n_2 n_3 \log(n_3) + (n_1 + n_2) n_3 r^2 + n_3 r^3)$.

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.

Anima Anandkumar, Prateek Jain, Yang Shi, and U. N. Niranjan. Tensor vs. matrix methods: Robust tensor decomposition under block sparse perturbations. In *International Conference on Artificial Intelligence and Statistics*, pages 268–276, 2016.

Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15(80):2773–2832, 2014.

HanQin Cai, Jian-Feng Cai, and Ke Wei. Accelerated alternating projections for robust principal component analysis. *Journal of Machine Learning Research*, 20(20):1–33, 2019.

HanQin Cai, Jialin Liu, and Wotao Yin. Learned robust PCA: A scalable deep unfolding approach for high-dimensional outlier detection. In *Advances in Neural Information Processing Systems*, pages 16977–16989, 2021.

HanQin Cai, Chandra Kundu, Jialin Liu, and Wotao Yin. Deeply learned robust matrix completion for large-scale low-rank data recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026. doi: 10.1109/TPAMI.2026.3659041.

Jian-Feng Cai, Jingyang Li, and Dong Xia. Generalized low-rank plus sparse tensor estimation by fast Riemannian optimization. *Journal of the American Statistical Association*, 118(544):2588–2604, 2022.

Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2): 572–596, 2011.

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.

Yeshwanth Cherapanamjeri, Kartik Gupta, and Prateek Jain. Nearly optimal robust matrix completion. In *International Conference on Machine Learning*, pages 797–805, 2017.

Yuejie Chi, Yue M. Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269, 2019.

Harry Dong, Megna Shah, Sean Donegan, and Yuejie Chi. Deep unfolded tensor robust PCA with self-supervised learning. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1–5, 2023a.

Harry Dong, Tian Tong, Cong Ma, and Yuejie Chi. Fast and provable tensor robust principal component analysis via scaled gradient descent. *Information and Inference: A Journal of the IMA*, 12(3):1716–1758, 2023b.

Derek Driggs, Stephen Becker, and Jordan Boyd-Graber. Tensor robust principal component analysis: Better recovery with atomic norm regularization. *arXiv preprint*, 2019. URL https://arxiv.org/abs/1901.10991.

Simon S. Du, Wei Hu, and Jason D. Lee. Algorithmic regularization in learning deep homogeneous models: Layers are automatically balanced. In *Advances in Neural Information Processing Systems*, pages 382–393, 2018.

Quanxue Gao, Pu Zhang, Wei Xia, Deyan Xie, Xinbo Gao, and Dacheng Tao. Enhanced tensor RPCA and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6):2133–2140, 2021.

Rong Ge, Chi Jin, and Yi Zheng. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In *International Conference on Machine Learning*, pages 1233–1242, 2017.

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.

Quanquan Gu, Zhaoran Wang, and Han Liu. Low-rank and sparse structure pursuit via alternating minimization. In *International Conference on Artificial Intelligence and Statistics*, pages 600–609, 2016.

Rungang Han, Rebecca Willett, and Anru R. Zhang. An optimal statistical and computational framework for generalized tensor estimation. *The Annals of Statistics*, 50(1):1–29, 2022.

Xixi Jia, Hailin Wang, Jiangjun Peng, Xiangchu Feng, and Deyu Meng. Preconditioning matters: Fast global convergence of non-convex matrix factorization via scaled gradient descent. In *Advances in Neural Information Processing Systems*, pages 76202–76213, 2023.

Eric Kernfeld, Misha Kilmer, and Shuchin Aeron. Tensor-tensor products with invertible linear transforms. *Linear Algebra and its Applications*, 485:545–570, 2015.

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.

Misha E. Kilmer, Lior Horesh, Haim Avron, and Elizabeth Newman. Tensor-tensor algebra for optimal representation and compression of multiway data. *Proceedings of the National Academy of Sciences*, 118(28), 2021.

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

Hao Kong, Canyi Lu, and Zhouchen Lin. Tensor Q-rank: new data dependent definition of tensor rank. *Machine Learning*, 110(7):1867–1900, 2021.

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.

Qiuwei Li, Zhihui Zhu, and Gongguo Tang. The non-convex geometry of low-rank matrix optimization. *Information and Inference: A Journal of the IMA*, 8(1):51–96, 2019.

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.

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

Cong Ma, Kaizheng Wang, Yuejie Chi, and Yuxin Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. *Foundations of Computational Mathematics*, 20(3): 451–632, 2020.

Bamdev Mishra and Rodolphe Sepulchre. Riemannian preconditioning. *SIAM Journal on Optimization*, 26(1):635–660, 2016.

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`.

Praneeth Netrapalli, Niranjan U N, Sujay Sanghavi, Animashree Anandkumar, and Prateek Jain. Non-convex robust PCA. In *Advances in Neural Information Processing Systems*, pages 1107–1115, 2014.

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

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.

Zhen Qin, Michael B. Wakin, and Zhihui Zhu. Guaranteed nonconvex factorization approach for tensor train recovery. *Journal of Machine Learning Research*, 25(383):1–48, 2024.

Haiquan Qiu, Yao Wang, Shaojie Tang, Deyu Meng, and Quanming Yao. Fast and provable nonconvex tensor RPCA. In *International Conference on Machine Learning*, pages 18211–18249, 2022.

Holger Rauhut, Reinhold Schneider, and Željka Stojanac. Low rank tensor recovery via iterative hard thresholding. *Linear Algebra and its Applications*, 523:220–262, 2017.

Benjamin Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12(104):3413–3430, 2011.

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, 2021a.

Tian Tong, Cong Ma, and Yuejie Chi. Low-rank matrix recovery with scaled subgradient methods: Fast and robust convergence without the condition number. *IEEE Transactions on Signal Processing*, 69:2396–2409, 2021b.

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.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12:389–434, 2012.

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

Antoine Vacavant, Thierry Chateau, Alexis Wilhelm, and Laurent Lequièvre. A benchmark dataset for outdoor foreground/background extraction. In *Computer Vision - ACCV 2012 Workshops*, pages 291–300, 2013.

Tong Wu. Robust data clustering with outliers via transformed tensor low-rank representation. In *International Conference on Artificial Intelligence and Statistics*, pages 1756–1764, 2024.

Tong Wu. A robust approach for image clustering with outliers via tensor low-rank representation. *Machine Learning*, 115(2), 2026.

Dong Xia and Ming Yuan. On polynomial time methods for exact low-rank tensor completion. *Foundations of Computational Mathematics*, 19(6):1265–1313, 2019.

Jing-Hua Yang, Xi-Le Zhao, Teng-Yu Ji, Tian-Hui Ma, and Ting-Zhu Huang. Low-rank tensor train for tensor robust principal component analysis. *Applied Mathematics and Computation*, 367:124783, 2020.

Tian Ye and Simon S. Du. Global convergence of gradient descent for asymmetric low-rank matrix factorization. In *Advances in Neural Information Processing Systems*, pages 1429–1439, 2021.

Xinyang Yi, Dohyung Park, Yudong Chen, and Constantine Caramanis. Fast algorithms for robust PCA via gradient descent. In *Advances in Neural Information Processing Systems*, pages 4152–4160, 2016.

Rose Yu, Dehua Cheng, and Yan Liu. Accelerated online low-rank tensor learning for multivariate spatio-temporal streams. In *International Conference on Machine Learning*, pages 238–247, 2015.

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.

Wen-Jun Zeng and Hing Cheung So. Outlier-robust matrix completion via ℓ_p-minimization. *IEEE Transactions on Signal Processing*, 66(5):1125–1140, 2018.

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*, pages 3842–3849, 2014.

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, Jiashi Feng, Zhouchen Lin, and Shuicheng Yan. Tensor low-rank representation for data recovery and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(5):1718–1732, 2021.